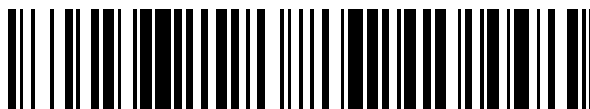


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 778 033**

51 Int. Cl.:

C12P 21/06

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.11.2013 PCT/US2013/070636**

87 Fecha y número de publicación internacional: **22.05.2014 WO14078819**

96 Fecha de presentación y número de la solicitud europea: **18.11.2013 E 13854876 (3)**

97 Fecha y número de publicación de la concesión europea: **19.02.2020 EP 2920319**

54 Título: **Enzimas específicas de sitio y métodos de uso**

30 Prioridad:

16.11.2012 US 201261727652 P
15.03.2013 US 201361802066 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
07.08.2020

73 Titular/es:

POSEIDA THERAPEUTICS, INC. (100.0%)
9390 Towne Centre Drive, Suite 200
San Diego, California 92121, US

72 Inventor/es:

OSTERTAG, ERIC M. y
YESHI, TSETEN

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 778 033 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Enzimas específicas de sitio y métodos de uso

5 Campo de la Invención

La presente invención se dirige, en parte, a proteínas de *Ralstonia*, ácidos nucleicos que codifican las mismas y kits que comprenden las mismas.

10 Antecedentes de la Invención

Los factores de transcripción con dominios de unión al ADN programables ofrecen un enfoque potencial para crear un circuito biológico exógeno en un sistema endógeno y crear proteínas de diseño que se unan a secuencias de ADN predeterminadas o ácidos nucleicos individuales. Recientemente se ha demostrado que las proteínas similares al activador de la transcripción (TAL) tienen dominios de unión al ADN modulares y predecibles, lo que permite de esta manera la creación *de novo* de factores de transcripción sintéticos que se unan a secuencias de ADN de interés y, si es conveniente, permiten que un segundo dominio presente en la proteína o polipéptido realice una actividad relacionada con el ADN. Las proteínas TAL se han derivado a partir del organismo *Xanthomonas*. Sin embargo, en la presente descripción se proporcionan proteínas o polipéptidos de *Ralstonia* que funcionan de manera similar a las proteínas TAL derivadas a partir de *Xanthomonas*.

El documento WO 2011/146121 describe polipéptidos que comprenden dominios de unión al ADN efectores-TAL. Jeffrey C Miller y otros. Nature Biotechnology, 29:2, pp.143-148 (2010) describen una estrategia para generar nucleasas que escinden secuencias genómicas únicas en células vivas, basadas en proteínas efectoras similares al activador de la transcripción (TALE) a partir de *Xanthomonas*. El documento US 2012/270273 describe composiciones, kits y métodos útiles en la construcción de polipéptidos efectores similares al activador de la transcripción de diseño (dTAL).

La invención se refiere a polipéptidos derivados a partir de secuencias de aminoácidos de *Ralstonia* o secuencias de aminoácidos relacionadas con estas, ácidos nucleicos que codifican las mismas y kits que comprenden las mismas.

30 Resumen de la invención

La invención se define por el conjunto de reivindicaciones adjuntas. Se basa en parte en el hecho de que los díresiduos variables de repetición (RVD) de efectores de *Ralstonia* corresponden a los nucleótidos en sus sitios objetivo de una manera directa y lineal, un RVD a un nucleótido, con alguna degeneración y sin dependencia aparente del contexto. Este hallazgo representa un mecanismo para el reconocimiento de ADN-proteína que permite la predicción de sitios objetivo para nuevos efectores de *Ralstonia* específicos de objetivo. Como se describe en la presente descripción, estas proteínas pueden ser útiles en investigación y biotecnología como parte de una proteína quimérica dirigida más grande con actividades accesorias relacionadas con ácidos nucleicos tales como la actividad nucleasa. Por ejemplo, en algunas modalidades, el polipéptido o pronucleasas que pueden facilitar la recombinación homóloga en la modificación genética del genoma (por ejemplo, para añadir o potenciar rasgos útiles para biocombustibles o productos biorenovables en plantas o bacterias, o en genomas animales). Además, estas proteínas pueden ser útiles como, por ejemplo, factores de transcripción, y especialmente para aplicaciones terapéuticas que requieren un nivel muy alto de especificidad, tales como los productos terapéuticos contra patógenos (por ejemplo, virus) como ejemplos no limitantes.

La invención proporciona un polipéptido que comprende al menos una secuencia de aminoácidos que tiene la secuencia de LSTEQVVAIASX₁X₂GGKQALEAVKAQLLVLRAPYE (SEQ ID NO:1); en donde, X₁ y X₂ comprenden díresiduos variables repetidos (RVD), y en donde el RVD consiste en SI, SN, SH, NP, NH, NT, NK, ND, HN, HY, HD, HH, RN, RS, NG o GS; y un dominio efector de ácido nucleico, en donde el dominio efector de ácido nucleico comprende una nucleasa, una nickasa, un activador de la transcripción, un represor transcripcional, una metiltransferasa, una desacetilasa o cualquier fragmento funcional de estos.

En algunas modalidades, el polipéptido comprende, además, una segunda secuencia de aminoácidos que tiene la secuencia de LSTEQVVAIASX₁X₂GGKQALEAVKAQLLVLRAPYE (SEQ ID NO: 1), en donde X₁X₂ comprende un díresiduo variable de repetición (RVD), en donde X₁ es un aminoácido de origen natural o no natural y en donde X₂ es un aminoácido de origen natural o no natural.

En otras modalidades, los polipéptidos comprenden, o no comprenden, la secuencia de aminoácidos de SEQ ID NO:1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 o la SEQ ID NO: 19.

En algunas modalidades, el dominio efector de ácido nucleico comprende una nucleasa. Por ejemplo, la nucleasa es una endonucleasa. En una modalidad, la endonucleasa comprende una endonucleasa Fok I. En una modalidad alternativa, la endonucleasa comprende una endonucleasa I-Sce I.

En otra modalidad, el dominio efector de ácido nucleico comprende una nucleasa de dedos de zinc.

La presente invención se refiere a secuencias de ácido nucleico que codifican cualquier polipéptido de la invención.

5 La presente invención proporciona, además, vectores que comprenden cualquiera de las secuencias de ácido nucleico de la invención. En algunas modalidades, el vector es un plásmido. En algunas modalidades, el vector es un retrovirus. En algunas modalidades, el retrovirus comprende repeticiones terminales largas, una señal de empaquetamiento *psi*, un sitio de clonación y una secuencia que codifica un marcador seleccionable.

10 La presente invención proporciona, además, células que comprenden cualquiera de los ácidos nucleicos o vectores de la invención. En algunas modalidades, la célula es un espermatozoide o un óvulo.

La presente invención proporciona, además, kits que comprenden: un vector que comprende un ácido nucleico que codifica cualquiera de los polipéptidos de la invención.

15 Breve descripción de los dibujos

La Figura 1 representa una secuencia consenso de una proteína de unión al ADN de *Xanthomonas* alineada mediante BLAST a secuencias de metiltransferasa de cepas bacterianas. Basado en la alineación de secuencia, se predice la función de unión al ADN de las secuencias.

La Figura 2 representa un gel que demuestra la funcionalidad RTN.

Descripción detallada

25 La invención se define por el conjunto de reivindicaciones adjuntas.

El polipéptido o las proteínas descritos en la presente descripción comprenden al menos un primer dominio y un segundo dominio, en donde el primer dominio comprende al menos una secuencia codificante para un elemento de reconocimiento de ácido nucleico y en donde el segundo dominio comprende al menos una secuencia codificante para un elemento efector de ácido nucleico. El polipéptido o las proteínas de la descripción comprenden al menos un primer dominio, en donde el primer dominio comprende al menos una secuencia codificante para un elemento de reconocimiento de ácido nucleico derivado de una secuencia de aminoácidos derivada a partir de *Ralstonia* o una variante de este. El término "RTN" se refiere a un polipéptido o proteína que comprende al menos un primer dominio, en donde el primer dominio comprende al menos una secuencia codificante para un elemento de reconocimiento de ácido nucleico derivado de una secuencia de aminoácidos derivada a partir de *Ralstonia*. En algunas modalidades, el término "RTN" se refiere a un polipéptido o proteína que comprende al menos un primer dominio y un segundo dominio, en donde el primer dominio comprende al menos una secuencia codificante para un elemento de reconocimiento de ácido nucleico derivado de una secuencia de aminoácidos derivada a partir de *Ralstonia* y el segundo dominio comprende un aminoácido que es una proteína efectora. En algunas modalidades, el término "RTN" se refiere a un polipéptido o proteína que comprende al menos un primer dominio y un segundo dominio, en donde el primer dominio comprende al menos una secuencia codificante para un elemento de reconocimiento de ácido nucleico derivado de una secuencia de aminoácidos derivada a partir de *Ralstonia* y el segundo dominio comprende un aminoácido que es una nucleasa. La especificidad de unión al ADN del RTN depende del número y el orden de los dominios de repetición en el dominio de unión al ADN. Las repeticiones generalmente se componen de aproximadamente 30 a aproximadamente 40 aminoácidos. En algunas modalidades, los dominios de repetición comprenden de aproximadamente 32 a aproximadamente 38 aminoácidos. En algunas modalidades, la especificidad de unión a nucleótidos se determina por los aminoácidos 12 y 13 de cada dominio de repetición. En algunas modalidades, los dominios de repetición comprenden de aproximadamente 33 a aproximadamente 37 aminoácidos. En algunas modalidades, los dominios de repetición comprenden de aproximadamente 34 a aproximadamente 35 aminoácidos. En algunas modalidades, los dominios de repetición comprenden de aproximadamente 33 a aproximadamente 36 aminoácidos. En algunas modalidades, los dominios de repetición comprenden de aproximadamente 33 a aproximadamente 35 aminoácidos. En algunas modalidades, los dominios de repetición consisten en 34 a 35 aminoácidos. En algunas modalidades, los dominios de repetición consisten en 33 a 35 aminoácidos. En algunas modalidades, los dominios de repetición consisten en 34 a 36 aminoácidos.

55 "Ácido nucleico" u "oligonucleótido" o "polinucleótido", como se usa en la presente, significa al menos dos nucleótidos unidos entre sí covalentemente. La representación de una cadena sencilla define, además, la secuencia de la cadena complementaria. Por lo tanto, un ácido nucleico abarca, además, la cadena complementaria de una cadena sencilla representada. Pueden usarse muchas variantes de un ácido nucleico para el mismo propósito que un ácido nucleico dado. Por lo tanto, un ácido nucleico abarca, además, ácidos nucleicos sustancialmente idénticos y sus complementos. Una cadena sencilla proporciona una sonda que puede hibridarse con una secuencia objetivo en condiciones de hibridación rigurosas. Por lo tanto, un ácido nucleico abarca, además, una sonda que hibrida en condiciones de hibridación rigurosas. Los ácidos nucleicos pueden ser monocatenarios o bicatenarios, o pueden contener tanto porciones de secuencia bicatenaria como monocatenaria. El ácido nucleico puede ser ADN, tanto genómico como ADNc, ARN o un híbrido, donde el ácido nucleico puede contener combinaciones de desoxirribosa y ribonucleótidos, y combinaciones de bases, lo que incluye uracilo, adenina, timina, citosina, guanina, inosina, xantina, hipoxantina, isocitosina e isoguanina. En algunas

modalidades, se sintetiza, comprende modificaciones de aminoácidos no naturales. Los ácidos nucleicos pueden obtenerse por métodos de síntesis química o por métodos recombinantes.

5 "Unido operativamente", como se usa en la presente, significa que la expresión de un gen está bajo el control de un promotor con el que se conecta espacialmente. Un promotor puede colocarse en dirección 5' (aguas arriba) o 3' (aguas abajo) de un gen bajo su control. La distancia entre el promotor y un gen puede ser aproximadamente la misma que la distancia entre ese promotor y el gen que controla en el gen a partir del que se deriva el promotor. Como se conoce en la técnica, la variación en esta distancia puede acomodarse sin pérdida de la función del promotor.

10 "Promotor", como se usa en la presente, significa una molécula sintética o de origen natural que es capaz de conferir, activar o potenciar la expresión de un ácido nucleico en una célula. Un promotor puede comprender una o más secuencias reguladoras de la transcripción específicas para potenciar aún más la expresión y/o alterar la expresión espacial y/o la expresión temporal de la misma. Un promotor puede comprender, además, elementos potenciadores o represores distales, que pueden localizarse hasta varios miles de pares de bases del sitio de inicio de la transcripción. Un promotor
15 puede derivarse de fuentes que incluyen virus, bacterias, hongos, plantas, insectos y animales. Un promotor puede regular la expresión de un componente génico constitutivamente o diferencialmente con respecto a la célula, el tejido u órgano en el que se produce la expresión o, con respecto a la etapa de desarrollo en la que se produce la expresión, o en respuesta a estímulos externos tales como estrés fisiológico, patógenos, iones metálicos, o agentes de inducción. Ejemplos representativos de promotores incluyen el promotor bacteriófago T7, el promotor bacteriófago T3, el promotor SP6, el promotor operador lac, el promotor tac, el promotor tardío SV40, el promotor temprano SV40, el promotor RSV-LTR, el promotor IE de CMV, el promotor temprano SV40 o el promotor tardío SV40 y el promotor IE de CMV,
20

"Sustancialmente complementaria", como se usa en la presente, significa que una primera secuencia es al menos 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 % o 99 % idéntica al complemento de una segunda secuencia a lo largo de una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 180, 270, 360, 450, 540 o más nucleótidos o aminoácidos, o que las dos secuencias hibridan en condiciones de hibridación rigurosas.
25

"Sustancialmente idéntica", como se usa en la presente, significa que una primera y una segunda secuencia son al menos 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 97 %, 98 % o 99 % idénticas a lo largo de una región de 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 180, 270, 360, 450, 540 o más nucleótidos o aminoácidos, o con respecto a los ácidos nucleicos, si la primera secuencia es sustancialmente complementaria al complemento de la segunda secuencia.
30

"Variante", usado en la presente descripción con respecto a un ácido nucleico significa (i) una porción o fragmento de una secuencia de nucleótidos referenciada; (ii) el complemento de una secuencia de nucleótidos referenciada o una porción de esta; (iii) un ácido nucleico que es sustancialmente idéntico a un ácido nucleico referenciado o al complemento de este; o (iv) un ácido nucleico que hibrida en condiciones rigurosas con el ácido nucleico referenciado, el complemento de este, o una secuencia sustancialmente idéntica al mismo.
35

40 Como se usa en la presente, "variante", con respecto a un péptido o polipéptido que difiere en la secuencia de aminoácidos mediante la inserción, delección o sustitución conservadora de aminoácidos, pero retiene al menos una actividad biológica. Variante puede significar, además, una proteína con una secuencia de aminoácidos que es sustancialmente idéntica a una proteína referenciada con una secuencia de aminoácidos que retiene al menos una actividad biológica. Una
45 sustitución conservadora de un aminoácido, es decir, reemplazar un aminoácido con un aminoácido diferente de propiedades similares (por ejemplo, hidrofiliicidad, grado y distribución de regiones cargadas) se reconoce en la técnica como que implica típicamente un cambio menor. Estos cambios menores pueden identificarse, en parte, al considerar el índice hidropático de los aminoácidos, como se entiende en la técnica. Kyte y otros, J. Mol. Biol. 157: 105-132 (1982). El índice hidropático de un aminoácido se basa en una consideración de su hidrofobicidad y su carga. Se conoce en la
50 técnica que los aminoácidos de índices hidropáticos similares pueden sustituirse y aún retener la función de la proteína. En un aspecto, los aminoácidos que tienen índices hidropáticos de ± 2 tienen sustituciones. La hidrofiliicidad de los aminoácidos puede usarse, además, para revelar sustituciones que resultarían en proteínas que retengan la función biológica. Una consideración de la hidrofiliicidad de los aminoácidos en el contexto de un péptido permite el cálculo de la mayor hidrofiliicidad promedio local de ese péptido, una medida útil que se ha informado que se correlaciona bien con la antigenicidad y la inmunogenicidad. Patente de Estados Unidos núm. 4,554,101.
55

La sustitución de aminoácidos que tienen valores de hidrofiliicidad similares puede resultar en que los péptidos retengan la actividad biológica, por ejemplo, la inmunogenicidad, como se entiende en la técnica. Las sustituciones pueden realizarse con aminoácidos que tienen valores de hidrofiliicidad dentro de ± 2 entre sí. Tanto el índice de hidrofobicidad como el valor de hidrofiliicidad de los aminoácidos se influyen por la cadena lateral particular de ese aminoácido. Coherentemente con esa observación, las sustituciones de aminoácidos que son compatibles con la función biológica se entiende que dependen de la similitud relativa de los aminoácidos, y particularmente de las cadenas laterales de esos aminoácidos, como lo revela la hidrofobicidad, hidrofiliicidad, carga, tamaño y otras propiedades.
60

65 "Vector", como se usa en la presente, significa una secuencia de ácido nucleico que contiene un origen de replicación. Un vector puede ser un vector viral, bacteriófago, cromosoma artificial bacteriano o cromosoma artificial de levadura. Un

vector puede ser un vector de ADN o ARN. Un vector puede ser un vector extracromosómico autorreplicante y, preferentemente, es un plásmido de ADN.

La descripción proporciona polipéptidos, proteínas y secuencias de ácidos nucleicos que codifican cualquiera de los polipéptidos o proteínas de la descripción. En algunas modalidades, la proteína comprende al menos un 75 % de identidad de secuencia con la SEQ ID NO:1. En algunas modalidades, la proteína comprende al menos un 75 % de identidad de secuencia con la SEQ ID NO:1 y comprende un dominio de unión a ácido nucleico en los aminoácidos 12º y 13º de la SEQ ID NO:1. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una secuencia de RVD seleccionada entre las siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS y GS. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una o una pluralidad de secuencias de RVD en cualquier combinación seleccionada entre las siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS, NG y GS; en donde SI, SN, SH, NP y NH se unen a cualquier base de ácido nucleico; en donde NT, NK y NN se unen a adenina; en donde ND, HN, HY, HD y HH se unen a adenina y/o guanina; en donde NG se une a timina; en donde RN, RS y GS se unen a guanina. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una o una pluralidad de secuencias de RVD en cualquier combinación seleccionada entre las siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS, NG y GS; en donde SI, SN, SH, NP y NH se unen a cualquier base de ácido nucleico; en donde NK se une a guanina y NN se une a adenina o guanina; en donde ND, HN, HY, HD y HH se unen a citosina; en donde NG se une a timina; en donde RN, RS y GS se unen a guanina. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una o una pluralidad de secuencias de RVD en cualquier combinación seleccionada entre las siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS, NG y GS; en donde SI se une a adenina; SN se une a guanina y/o adenina, SH, NP y NH se unen a cualquier base de ácido nucleico; en donde NK se une a guanina; y NN se une a adenina y/o guanina; en donde ND se une a citosina, HN se une a guanina, HY, HD y HH se une a citosina; en donde NG se une a timina; en donde RN se une a guanina y/o adenina; en donde RS y GS se unen a guanina. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una o una pluralidad de secuencias de RVD en cualquier combinación en donde al menos una de las secuencias de RVD es NP, ND o HN; y en donde NP se une a citosina, adenina y guanina; en donde ND se une a citosina; y en donde HN se une a adenina y/o guanina.

En algunas modalidades de la descripción, la proteína comprende al menos el 75 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 80 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 85 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 90 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 91 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 92 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 93 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 94 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 95 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 96 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 97 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 98 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13. En algunas modalidades, la proteína comprende al menos un 99 % de identidad de secuencia con la SEQ ID NO:1, y comprende una sustitución conservadora de aminoácidos en al menos una de las siguientes posiciones de aminoácidos en la SEQ ID NO:1: posición 12 y posición 13.

En algunas modalidades, la proteína comprende al menos un 75 % de identidad de secuencia con la SEQ ID NO:1.

En algunas modalidades, la proteína (como ácido nucleico, como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 80 % de identidad de secuencia con la SEQ ID NO:2, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:2. En algunas modalidades, la proteína (como ácido nucleico, como ácido nucleico en un vector, o como proteína recombinante purificada) comprende

al menos un 85 % de identidad de secuencia con la SEQ ID NO:2, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:2. En algunas modalidades, la proteína (como ácido nucleico, como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 90 % de identidad de secuencia con la SEQ ID NO:2, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:2. En algunas modalidades, la proteína (como ácido nucleico, como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 95 % de identidad de secuencia con la SEQ ID NO:2, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:2. En algunas modalidades, la proteína (como ácido nucleico, como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 99 % de identidad de secuencia con la SEQ ID NO:2, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:2.

En algunas modalidades, la proteína (codificada por un ácido nucleico o como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 80 % de identidad de secuencia con la SEQ ID NO:3, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:3. En algunas modalidades, la proteína (codificada por un ácido nucleico o como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 85 % de identidad de secuencia con la SEQ ID NO:3, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:3. En algunas modalidades, la proteína (codificada por un ácido nucleico o como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 90 % de identidad de secuencia con la SEQ ID NO:3, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:3. En algunas modalidades, la proteína (codificada por un ácido nucleico o como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 95 % de identidad de secuencia con la SEQ ID NO:3, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:3. En algunas modalidades, la proteína (codificada por un ácido nucleico o como ácido nucleico en un vector, o como proteína recombinante purificada) comprende al menos un 99 % de identidad de secuencia con la SEQ ID NO:3, y comprende al menos una de las sustituciones de aminoácidos mencionadas anteriormente en la SEQ ID NO:3.

[illegible]

cualquiera o más de la SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19.

Como se usa en la presente, la "identidad de secuencia" se determina mediante el uso del programa de búsqueda BLAST ejecutable independiente para el alineamiento de dos secuencias (b12seq), que puede recuperarse del sitio ftp del National Center for Biotechnology Information (NCBI), mediante el uso de los parámetros predeterminados (Tatusova y Madden, FEMS Microbiol Lett., 1999, 174, 247-250). "Idéntico" o "identidad", como se usa en la presente en el contexto de dos o más ácidos nucleicos o secuencias de polipéptidos, significa que las secuencias tienen un porcentaje especificado de residuos que son iguales a lo largo de una región especificada. El porcentaje puede calcularse al alinear de manera óptima las dos secuencias, comparar las dos secuencias a lo largo de la región especificada, determinar el número de posiciones en las que se presente un residuo idéntico en ambas secuencias para dar el número de posiciones coincidentes, dividir el número de posiciones coincidentes por el número total de posiciones en la región especificada, y multiplicar los resultados por 100 para obtener el porcentaje de identidad de secuencia. En los casos en que las dos secuencias son de diferentes longitudes o el alineamiento produce uno o más extremos escalonados y la región de comparación especificada incluye solo una secuencia, los residuos de secuencia única se incluyen en el denominador pero no en el numerador del cálculo. Cuando se comparan ADN y ARN, la timina (T) y el uracilo (U) pueden considerarse equivalentes. La identidad puede realizarse manualmente o mediante el uso de un algoritmo de secuencia de computadora tal como BLAST o BLAST 2.0.

Como se usa en la presente, las sustituciones de aminoácidos "conservadoras" pueden definirse como se expone en las Tablas A, B o C más abajo. En algunas modalidades, los polipéptidos de fusión y/o los ácidos nucleicos que codifican dichos polipéptidos de fusión incluyen sustituciones conservadoras que se han introducido mediante la modificación de los polinucleótidos que codifican dichos polipéptidos. Los aminoácidos pueden clasificarse de acuerdo con las propiedades físicas y la contribución a la estructura proteica secundaria y terciaria. Una sustitución conservadora se reconoce en la técnica como una sustitución de un aminoácido por otro aminoácido que tiene propiedades similares. Sustituciones conservadoras ilustrativas se exponen en la Tabla A.

Tabla A -- Sustituciones conservadoras I

Características de la cadena lateral	Aminoácido
Alifática	
No polar	G A P I L V F
Polar - sin carga	C S T M N Q
Polar - cargada	D E K R
Aromática	H F W Y
Otra	N Q D E

Alternativamente, los aminoácidos conservadores pueden agruparse como se describe en Lehninger, (Biochemistry, Segunda Edición; Worth Publishers, Inc. NY, NY (1975), pp. 71-77) como se expone en la Tabla B.

Tabla B -- Sustituciones conservadoras II

Característica de la cadena lateral	Aminoácido
No polar (hidrófobo)	
Alifático:	A L I V P.
Aromático:	F W Y
Que contiene azufre:	M
Indeterminado:	G Y
Polar-sin carga	
Hidroxilo:	S T Y
Amidas:	N Q
Sulfhidrido:	C
Indeterminado:	G Y
Cargado positivamente (básico):	K R H
Cargado negativamente (ácido):	D E

Alternativamente, sustituciones conservadoras ilustrativas se exponen en la Tabla C.

Tabla C -- Sustituciones conservadoras III

	Residuo original	Sustitución ilustrativa
5	Ala (A)	Val Leu Ile Met
	Arg (R)	Lys His
	Asn (N)	Gln
10	Asp (D)	Glu
	Cys (C)	Ser Thr
	Gln (Q)	Asn
15	Glu (E)	Asp
	Gly (G)	Ala Val Leu Pro
	His (H)	Lys Arg
20	Ile (I)	Leu Val Met Ala Phe
	Leu (L)	Ile Val Met Ala Phe
	Lys (K)	Arg His
25	Met (M)	Leu Ile Val Ala
	Phe (F)	Trp Tyr Ile
	Pro (P)	Gly Ala Val Leu Ile
30	Ser (S)	Thr
	Thr (T)	Ser
	Trp (W)	Tyr Phe Ile
35	Tyr (Y)	Trp Phe Thr Ser
	Val (V)	Ile Leu Met Ala

Debe entenderse que los polipéptidos descritos en la presente descripción pretenden incluir polipéptidos que portan una o más inserciones, deleciones o sustituciones, o cualquiera de sus combinaciones, de residuos de aminoácidos, así como también modificaciones distintas a las inserciones, deleciones o sustituciones de residuos de aminoácidos. En algunas modalidades, los polipéptidos o ácidos nucleicos descritos en la presente descripción contienen una o más sustituciones conservadoras. En algunas modalidades, los polipéptidos o ácidos nucleicos descritos en la presente descripción contienen más de una sustitución conservadora.

Como se usa en la presente, "más de una" de las sustituciones de aminoácidos mencionadas anteriormente significa 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más de las sustituciones de aminoácidos mencionadas. En algunas modalidades, "más de una" significa 2, 3, 4 o 5 de las sustituciones de aminoácidos mencionadas. En algunas modalidades, "más de una" significa 2, 3 o 4 de las sustituciones de aminoácidos mencionadas. En algunas modalidades, "más de una" significa 2 o 3 de las sustituciones de aminoácidos mencionadas. En algunas modalidades, "más de una" significa 2 de las sustituciones de aminoácidos mencionadas.

Como se usa en la presente, "endógena" se refiere a una secuencia de ácido nucleico o proteína asociada naturalmente con un gen objetivo o una célula huésped en la que se introduce.

Como se usa en la presente, "exógena" se refiere a una secuencia de ácido nucleico o proteína no asociada naturalmente con un gen objetivo o una célula huésped en la que se introduce, lo que incluye múltiples copias de origen no natural de un ácido nucleico de origen natural, por ejemplo, secuencia de ADN, o secuencia de ácido nucleico de origen natural ubicada en una ubicación del genoma de origen no natural.

Como se usa en la presente, "planta modificada genéticamente (o planta transgénica)" se refiere a una planta que comprende dentro de su genoma un polinucleótido exógeno. Generalmente, y preferentemente, el polinucleótido exógeno se integra de manera estable dentro del genoma de manera que el polinucleótido se transmite a las generaciones sucesivas. El polinucleótido exógeno puede integrarse en el genoma solo o como parte de un casete de expresión recombinante. "Transgénico" se usa en la presente descripción para incluir cualquier célula, línea celular, callo, tejido, parte de planta o planta, cuyo genotipo se ha alterado por la presencia de ácido nucleico exógeno, lo que incluye aquellas transgénicas inicialmente ya alteradas, así como también las creadas por cruces sexuales o propagación asexual a partir

del transgénico inicial. El término "transgénico", como se usa en la presente, no abarca la alteración del genoma (cromosómica o extracromosómica) por métodos convencionales de fitomejoramiento o por eventos de origen natural tales como fertilización cruzada aleatoria, infección viral no recombinante, transformación bacteriana no recombinante, transposición no recombinante o mutación espontánea.

El término "modificar", como se usa en la presente, pretende significar que la secuencia se considera modificada simplemente por la unión del polipéptido. No se pretende sugerir que se cambió la secuencia de nucleótidos, aunque dichos cambios (y otros) podrían producirse después de la unión del polipéptido al ácido nucleico de interés. En algunas modalidades, la secuencia de ácido nucleico es ADN. La modificación del ácido nucleico de interés (en el sentido de la unión al mismo por un polipéptido modificado para contener unidades de repetición modulares) podría detectarse en cualquiera de una serie de métodos (por ejemplo, ensayos de cambio de movilidad en gel, uso de polipéptidos marcados-los marcadores podrían incluir marcadores radiactivos, fluorescentes, enzimas o biotina/estreptavidina). La modificación de la secuencia de ácido nucleico de interés (y la detección de esta) puede ser todo lo que se requiere (por ejemplo, en el diagnóstico de la enfermedad). Sin embargo, convenientemente, se realiza un procesamiento adicional de la muestra. Convenientemente, el polipéptido (y las secuencias de ácidos nucleicos específicamente unidas al mismo) se separa del resto de la muestra. Ventajosamente, el complejo polipéptido-ADN se une a un soporte en fase sólida, para facilitar dicha separación. Por ejemplo, el polipéptido puede presentarse en una matriz de gel de acrilamida o agarosa o, con mayor preferencia, se inmoviliza en la superficie de una membrana o en los pocillos de una placa de microtitulación.

En algunas modalidades, las proteínas de fusión descritas en la presente descripción comprenden al menos dos dominios, en donde el primer dominio es un elemento de unión al ADN de *Ralstonia* y el segundo dominio es una metilasa.

Las secuencias de ADN descritas en la presente descripción pueden proporcionarse en casetes de expresión para la expresión en cualquier célula y/u organismo procariota o eucariota de interés lo que incluye, pero no se limita a, bacterias, hongos, algas, plantas y animales. El casete incluirá secuencias reguladoras 5' y 3' unidas operativamente a la secuencia de ADN. "Unido operativamente" pretende significar una unión funcional entre dos o más elementos. Por ejemplo, una unión operativa entre un polinucleótido o gen de interés y una secuencia reguladora (es decir, un promotor) es un enlace funcional que permite la expresión del polinucleótido de interés. Los elementos unidos operativamente pueden ser contiguos o no contiguos. Cuando se usa para referirse a la unión de dos regiones de codificación de proteínas, por unidas operativamente se pretende que las regiones de codificación están en el mismo marco de lectura. El casete puede contener adicionalmente al menos un gen adicional a cotransformarse en el organismo. Alternativamente, el(los) gen(es) adicional(es) puede(n) proporcionarse en casetes de expresión múltiple. Dicho casete de expresión se proporciona con una pluralidad de sitios de restricción y/o sitios de recombinación para que la inserción de la secuencia de ADN esté bajo la regulación transcripcional de las regiones reguladoras. El casete de expresión puede contener adicionalmente genes marcadores seleccionables.

El casete de expresión incluirá en la dirección de transcripción 5'-3', una región de inicio de transcripción y traducción (es decir, un promotor), una secuencia de ADN y una región de terminación de transcripción y traducción (es decir, región de terminación) funcional en plantas u otro organismo o célula huésped no humana. Las regiones reguladoras (es decir, los promotores, las regiones reguladoras de la transcripción y las regiones de terminación de la traducción) y/o la secuencia de ADN pueden ser nativas/análogas con respecto a la célula huésped o entre sí. Alternativamente, las regiones reguladoras y/o la secuencia de ADN pueden ser heterólogas con respecto a la célula huésped o entre sí. Como se usa en la presente, "heteróloga" en referencia a una secuencia es una secuencia que se origina a partir de una especie extraña, o, si es de la misma especie, se modifica sustancialmente de su forma nativa en composición y/o locus genómico mediante intervención humana deliberada. Por ejemplo, un promotor unido operativamente a un polinucleótido heterólogo es de una especie diferente de la especie de la que se deriva el polinucleótido, o, si es de la misma especie/análoga, uno o ambos se modifican sustancialmente de su forma original y/o locus genómico, o el promotor no es el promotor nativo para el polinucleótido unido operativamente. Como se usa en la presente, un gen quimérico comprende una secuencia codificante unida operativamente a una región de inicio de la transcripción que es heteróloga con respecto a la secuencia codificante.

La región de terminación puede ser nativa con la región de inicio de la transcripción, puede ser nativa con la secuencia de ADN unida operativamente de interés, puede ser nativa con el huésped, o puede derivarse de otra fuente (es decir, extraña o heteróloga) con respecto al promotor, la secuencia de ADN de interés, la planta huésped, o cualquiera de sus combinaciones. Regiones de terminación convenientes para su uso en plantas están disponibles del plásmido Ti de *A. tumefaciens*, tal como las regiones de terminación de la octopina sintasa y nopalina sintasa. Ver, además, Guerineau y otros (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon y otros. (1991) *Genes Dev.* 5:141-149; Mogen y otros (1990) *Plant Cell* 2:1261-1272; Munroe y otros (1990) *Gene* 91:151-158; Ballas y otros (1989) *Nucleic Acids Res.* 17:7891-7903; y Joshi y otros (1987) *Nucleic Acids Res.* 15:9627-9639.

Cuando sea apropiado, los polinucleótidos pueden optimizarse para aumentar la expresión en un organismo transformado. Es decir, los polinucleótidos pueden sintetizarse mediante el uso de codones preferidos por el huésped para una expresión mejorada. Ver, por ejemplo, Campbell y Gown (1990) *Plant Physiol.* 92:1-11 para una discusión sobre el uso de codones preferidos por el huésped. Los métodos están disponibles en la técnica para sintetizar genes preferidos por el huésped, particularmente genes preferidos por las plantas. Ver, por ejemplo, las Patentes de Estados Unidos Núms. 5,380,831, y 5,436,391, y Murray y otros (1989) *Nucleic Acids Res.* 17:477-498.

Se conocen modificaciones de secuencia adicionales para mejorar la expresión génica en un huésped celular. Estas incluyen la eliminación de secuencias que codifican señales de poliadenilación espurias, señales de sitio de empalme exón-intrón, repeticiones similares a transposones y otras dichas secuencias bien caracterizadas que pueden ser perjudiciales para la expresión génica. El contenido de G-C de la secuencia puede ajustarse a niveles promedio para un huésped celular dado, según se calcula por referencia con respecto a genes conocidos expresados en la célula huésped. Cuando es posible, la secuencia se modifica para evitar las estructuras de ARNm secundarias de horquilla predichas.

Los casetes de expresión pueden contener adicionalmente secuencias líderes 5'. Dichas secuencias líderes pueden actuar para potenciar la traducción. Los líderes de traducción se conocen en la técnica e incluyen: líderes de picornavirus, por ejemplo, líder de EMCV (región no codificante 5' de encefalomiocarditis) (Elroy-Stein y otros (1989) Proc. Natl. Acad. Sci. Estados Unidos 86:6126-6130); líderes de polivirus, por ejemplo, líder de TEV (virus del grabado del tabaco) (Gallie y otros (1995) Gene 165(2):233-238), líder del MDMV (Virus del mosaico enano del maíz) (Virology 154:9-20) y la proteína de unión a la cadena pesada de inmunoglobulina humana (BiP) (Macejak y otros (1991) Nature 353:90-94); líder no traducido del ARNm de la proteína de la cubierta del virus del mosaico de la alfalfa (AMV RNA 4) (Tabling y otros (1987) Nature 325:622-625); líder del virus del mosaico del tabaco (TMV) (Gallie y otros (1989) en Molecular Biology of RNA, ed. Cech (Liss, Nueva York), pp. 237-256); y líder del virus del moteado clorótico del maíz (MCMV) (Lommel y otros (1991) Virology 81:382-385). Ver, además, Della-Cioppa y otros (1987) Plant Physiol. 84:965-968.

Al preparar el casete de expresión, los diversos fragmentos de ADN pueden manipularse para proporcionar las secuencias de ADN en la orientación apropiada y, según sea apropiado, en el marco de lectura apropiado. Con este fin, pueden emplearse adaptadores o enlazadores para unir los fragmentos de ADN o pueden involucrarse otras manipulaciones para proporcionar sitios de restricción convenientes, eliminación de ADN superfluo, eliminación de sitios de restricción, o similares. Para este propósito, puede involucrarse mutagénesis *in vitro*, reparación con cebadores, restricción, hibridación, resustituciones, por ejemplo, transiciones y transversiones.

La descripción abarca composiciones de polinucleótidos o proteínas aisladas o sustancialmente purificadas. Un polinucleótido o proteína "aislada" o "purificada", o una porción biológicamente activa de esta, está sustancial o esencialmente libre de componentes que normalmente acompañan o interactúan con el polinucleótido o proteína como se encuentra en su entorno de origen natural. De esta forma, un polinucleótido o proteína aislada o purificada está sustancialmente libre de otro material celular, o medio de cultivo cuando se produce mediante técnicas recombinantes, o sustancialmente libre de precursores químicos u otras sustancias químicas cuando se sintetiza químicamente. Óptimamente, un polinucleótido "aislado" está libre de secuencias (óptimamente secuencias que codifican proteínas) que flanquean naturalmente al polinucleótido (por ejemplo, secuencias ubicadas en los extremos 5' y 3' del polinucleótido) en el ADN genómico del organismo del cual se deriva el polinucleótido. Por ejemplo, en diversas modalidades, el polinucleótido aislado puede contener menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencia de nucleótidos que flanquean naturalmente el polinucleótido en el ADN genómico de la célula del cual se deriva el polinucleótido. Una proteína que está sustancialmente libre de material celular incluye preparaciones de proteína que tienen menos de aproximadamente 30 %, 20 %, 10 %, 5 % o 1 % (en peso seco) de proteína contaminante. Cuando la proteína o porción biológicamente activa de esta se produce de forma recombinante, el medio de cultivo óptimo representa menos del 30 %, 20 %, 10 %, 5 % o 1 % (en peso seco) de precursores químicos o sustancias químicas que no son las proteínas de interés.

Los fragmentos y variantes de las secuencias de ADN descritas y las proteínas codificadas de esta manera también se describen en la presente descripción. Por "fragmento" se entiende una porción de la secuencia de ADN o una porción de la secuencia de aminoácidos y, por lo tanto, la proteína codificada de esta manera. Los fragmentos de una secuencia de ADN que comprende secuencias codificantes pueden codificar fragmentos de proteínas que retienen la actividad biológica de la proteína nativa y, por lo tanto, el reconocimiento del ADN o la actividad de unión a una secuencia de ADN objetivo como se describe en la presente descripción. Alternativamente, los fragmentos de una secuencia de ADN que son útiles como sondas de hibridación generalmente no codifican proteínas que retienen la actividad biológica o no retienen la actividad del promotor. Por lo tanto, los fragmentos de una secuencia de ADN pueden variar de al menos aproximadamente 20 nucleótidos, aproximadamente 50 nucleótidos, aproximadamente 100 nucleótidos y hasta el polinucleótido de longitud completa.

En algunas modalidades, la proteína comprende la SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 y/o SEQ ID NO:19. En algunas modalidades, una cualquiera de las secuencias de polipéptidos SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 y/o SEQ ID NO:19 se repite al menos una vez. En algunas modalidades, el polipéptido no comprende ninguna de las secuencias de la Tabla 1. En algunas modalidades, el polipéptido comprende una sola secuencia en la Tabla 1 pero no comprende al menos una o más de las secuencias alternativas en la Tabla 1. En algunas modalidades, la secuencia alternativa comprende la SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19.

- La descripción proporciona, además, ácidos nucleicos que codifican una cualquiera de las proteínas polipeptídicas descritas en la presente descripción. De esta forma, la descripción proporciona ácidos nucleicos que codifican una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1.
- 5 En algunas modalidades, el ácido nucleico codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1, donde X_1X_2 son cualquier combinación de aminoácidos de origen natural o no natural.
- 10 Dada la redundancia en el código genético, un experto en la técnica podría generar numerosas secuencias de nucleótidos que codifican cualquier proteína en particular. Todas estas secuencias de nucleótidos se contemplan en la presente descripción.
- 15 La descripción proporciona, además, vectores que comprenden cualquiera de los ácidos nucleicos mencionados anteriormente. Por lo tanto, la descripción proporciona vectores que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1. La descripción proporciona vectores que comprenden un ácido nucleico que codifica al menos una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1, y comprende al menos una secuencia de RVD en sus aminoácidos 12º y 13º de la SEQ ID NO:1 seleccionados entre los siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS y GS. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una o una pluralidad de secuencias de RVD en cualquier combinación seleccionada entre las siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS, NG y GS; en donde SI, SN, SH, NP y NH se unen a cualquier base de ácido nucleico; en donde NT, NK y NN se unen a adenina; en donde ND, HN, HY, HD y HH se unen a adenina y/o guanina; en donde NG se une a timina; en donde RN, RS y GS se unen a guanina. En algunas modalidades, NK se une a guanina, NG se une a timina, NN se une a guanina o adenina, y/o HD se une a citosina. En algunas modalidades, SI se une a adenina, SN se une a guanina o adenina, ND se une a citosina, HN se une a guanina y/o RN se une a guanina o adenina.
- 20 SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS y GS. En algunas modalidades, las proteínas o polipéptidos comprenden al menos una o una pluralidad de secuencias de RVD en cualquier combinación seleccionada entre las siguientes: SI, SN, SH, NP, NH, NT, NK, NN, ND, HN, HY, HD, HH, RN, RS, NG y GS; en donde SI, SN, SH, NP y NH se unen a cualquier base de ácido nucleico; en donde NT, NK y NN se unen a adenina; en donde ND, HN, HY, HD y HH se unen a adenina y/o guanina; en donde NG se une a timina; en donde RN, RS y GS se unen a guanina. En algunas modalidades, NK se une a guanina, NG se une a timina, NN se une a guanina o adenina, y/o HD se une a citosina. En algunas modalidades, SI se une a adenina, SN se une a guanina o adenina, ND se une a citosina, HN se une a guanina y/o RN se une a guanina o adenina.
- 25 En algunas modalidades, el polipéptido comprende al menos un primer y un segundo dominio, en donde el primer dominio comprende al menos un monómero de polipéptido que comprende una sola secuencia de RVD descrita anteriormente. En algunas modalidades, el primer dominio comprende 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más monómeros, en donde cada monómero comprende un único dominio de unión a ácido nucleico que consiste en dos aminoácidos, o RVD. En algunas modalidades, el primer dominio comprende al menos dos monómeros en donde cada monómero se separa por un espaciador de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 o más aminoácidos.
- 30 En algunas modalidades, el segundo dominio comprende otra función complementaria con respecto a la unión a ácido nucleico conferida por la presencia del primer dominio. En algunas modalidades, el segundo dominio es una nucleasa o fragmento funcional de una nucleasa. En algunas modalidades, el segundo dominio es una endonucleasa o fragmento funcional de una endonucleasa. En algunas modalidades, el segundo dominio es una nickasa o fragmento funcional de una nickasa. En algunas modalidades, el segundo dominio es un represor o fragmento funcional de un represor. En algunas modalidades, el segundo dominio es un activador de la transcripción o fragmento funcional de un activador de la transcripción.
- 35 En algunas modalidades, el vector comprende un ácido nucleico que comprende una secuencia que codifica uno o más de cada polipéptido descrito en la presente descripción. En algunas modalidades, el vector comprende un ácido nucleico que comprende una secuencia que codifica uno o más de cada polipéptido elegido entre: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19.
- 40 En algunas modalidades, el vector es un plásmido. En otras modalidades, el vector es un retrovirus. En algunas modalidades, el vector es una molécula de ADN lineal. En algunas modalidades, el retrovirus comprende repeticiones terminales largas, una señal de empaquetamiento *psi*, un sitio de clonación y una secuencia que codifica un marcador seleccionable. En algunas modalidades, el vector es un vector viral, tal como pLXIN (Clontech).
- 45 La descripción proporciona, además, células u organismos que comprenden cualquiera de los ácidos nucleicos mencionados anteriormente. De esta forma, la descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos en cualquier combinación: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera
- 50 En algunas modalidades, el vector es un plásmido. En otras modalidades, el vector es un retrovirus. En algunas modalidades, el vector es una molécula de ADN lineal. En algunas modalidades, el retrovirus comprende repeticiones terminales largas, una señal de empaquetamiento *psi*, un sitio de clonación y una secuencia que codifica un marcador seleccionable. En algunas modalidades, el vector es un vector viral, tal como pLXIN (Clontech).
- 55 La descripción proporciona, además, células u organismos que comprenden cualquiera de los ácidos nucleicos mencionados anteriormente. De esta forma, la descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos en cualquier combinación: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera
- 60 La descripción proporciona, además, células u organismos que comprenden cualquiera de los ácidos nucleicos mencionados anteriormente. De esta forma, la descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos en cualquier combinación: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera
- 65 La descripción proporciona, además, células u organismos que comprenden cualquiera de los ácidos nucleicos mencionados anteriormente. De esta forma, la descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos en cualquier combinación: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. La descripción proporciona células u organismos que comprenden un ácido nucleico que codifica una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera

o más de los siguientes polipéptidos en cualquier combinación: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19; y comprende al menos una mutación en al menos uno de los dominios RVD en la posición 12 y 13 de cualquiera de las SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. La descripción proporciona células u organismos que comprenden un ácido nucleico mutado por contacto con una proteína que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos en cualquier combinación: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19; y comprende al menos una mutación en al menos uno de los dominios RVD en la posición 12 y 13 de cualquiera de las SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19.

En algunas modalidades, las células u organismos comprenden un ácido nucleico que codifica una proteína que comprende una secuencia de ácido nucleico con al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con la SEQ ID NO:1. En algunas modalidades, el polipéptido o proteína comprende múltiples dominios de repetición, en donde al menos un dominio de repetición comprende una secuencia de ácido nucleico con al menos 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) de identidad de secuencia con respecto a la SEQ ID NO:1-19 o una variante de este que tiene una identidad de secuencia del 75 % (u 80 %, 85 %, 90 %, 95 % o 99 %) con respecto a la SEQ ID NO:1-19.

En algunas modalidades, la célula comprende cualquiera de los vectores o secuencias de ácido nucleico mencionados anteriormente.

En un aspecto de la descripción, los polipéptidos comprenden subunidades monoméricas, en donde al menos una subunidad monomérica comprende al menos una secuencia de aminoácidos derivada de *Ralstonia* que comprende un elemento de reconocimiento de nucleótidos. En algunas modalidades, los polipéptidos comprenden subunidades monoméricas, en donde al menos una subunidad monomérica comprende al menos una secuencia de aminoácidos derivada de *Ralstonia* que comprende un elemento de reconocimiento de nucleótidos que comprende al menos 75 % (u 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. En algunas modalidades, los polipéptidos comprenden subunidades monoméricas, en donde al menos una subunidad monomérica comprende al menos una secuencia de aminoácidos derivada de *Ralstonia* que comprende un elemento de reconocimiento de nucleótidos que consiste en al menos 75 % (u 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 %) de identidad de secuencia con uno cualquiera o más de los siguientes polipéptidos: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19. En algunas modalidades, las moléculas de ácido nucleico de la descripción comprenden una secuencia de ácido nucleico que codifica una o más subunidades monoméricas, en donde la una o más subunidades monoméricas, cuando se codifican, comprenden dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, diez o más, once o más, o doce o más monómeros secuenciales, cada monómero comprende al menos una secuencia de aminoácidos derivada de *Ralstonia* que comprende un elemento de reconocimiento de nucleótidos. En algunas modalidades, las moléculas de ácido nucleico comprenden una secuencia de ácido nucleico que codifica una o más subunidades de monómero, en donde la una o más subunidades de monómero, cuando se codifican, comprenden dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, diez o más, once o más, o doce o más monómeros secuenciales, cada monómero comprende al menos una secuencia de aminoácidos derivada de *Ralstonia* que comprende un elemento de reconocimiento de nucleótidos que consiste en al menos 75 % (u 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 %) de identidad de secuencia con respecto a uno o más de los siguientes polipéptidos: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19 o cualquier variante o análogo de estas. En algunas modalidades, la proteína comprende una o más secuencias de ácido nucleico que codifican una o más secuencias de polipéptidos que comprenden un elemento de reconocimiento de ácido nucleico derivado de *Ralstonia* y una o más secuencias de ácido nucleico que codifican una o más secuencias de polipéptidos que comprenden un elemento de reconocimiento de ácido nucleico derivado de *Xanthomonas*. En algunas modalidades, la proteína de fusión comprende subunidades monoméricas sucesivas variables independientemente que son elementos de reconocimiento de ADN, en donde al menos una o más de las subunidades monoméricas se deriva de una secuencia de *Ralstonia* descrita en la presente descripción. En algunas modalidades, la descripción se refiere a una proteína de fusión que comprende subunidades monoméricas sucesivas variables independientemente que son elementos de reconocimiento de ADN, en donde al menos una o más de las subunidades monoméricas se deriva de una secuencia de *Ralstonia* descrita en la Tabla 1. En algunas modalidades, la descripción se refiere a una proteína de fusión que comprende subunidades

[illegible]

al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % por ciento homóloga a estas, en donde cada monómero se une a un nucleótido de una secuencia objetivo de ADN en presencia de una secuencia de ácido nucleico, y en donde la proteína de fusión comprende, además, al menos una secuencia de polipéptidos que es una proteína/polipéptido efectora y comprende, además, al menos una secuencia de polipéptidos que tiene cualquier función de proteína efectora descrita. En algunas modalidades, la descripción se refiere a una proteína de fusión que comprende subunidades monoméricas sucesivas que son elementos de reconocimiento de ADN, en donde la proteína de fusión comprende cualquier combinación de polipéptidos sucesivos elegidos de al menos una de las siguientes secuencias de polipéptidos: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19 o cualquier variante o análogo de estas que sea al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % por ciento homóloga a estas, en donde cada monómero se une a un nucleótido de una secuencia objetivo de ADN en presencia de una secuencia de ácido nucleico, y en donde la proteína de fusión comprende, además, al menos dos secuencias de polipéptidos que comprenden una función de proteína/polipéptido efectora o que son proteínas efectoras o variantes de estas. En algunas modalidades, la descripción se refiere a una proteína de fusión que comprende subunidades monoméricas sucesivas que son elementos de reconocimiento de ADN, en donde la proteína de fusión comprende cualquier combinación de polipéptidos sucesivos elegidos de al menos una de las siguientes secuencias de polipéptidos: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19 o cualquier variante o análogo de estas que sea al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % por ciento homóloga a estas, en donde cada monómero se une a un nucleótido de una secuencia objetivo de ADN en presencia de una secuencia de ácido nucleico, y en donde la proteína de fusión comprende, además, al menos tres o más secuencias de polipéptidos que comprenden una función proteína/efectora. En algunas modalidades, la descripción se refiere a una proteína de fusión que comprende subunidades monoméricas sucesivas que son elementos de reconocimiento de ADN, en donde la proteína de fusión comprende cualquier combinación de polipéptidos sucesivos elegidos de al menos una de las siguientes secuencias de polipéptidos: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18 o SEQ ID NO:19 o cualquier variante o análogo de estas que sea al menos 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 % o 99 % por ciento homóloga a estas, en donde cada monómero se une a un nucleótido de una secuencia objetivo de ADN en presencia de una secuencia de ácido nucleico, y en donde la proteína de fusión comprende, además, al menos cuatro o más secuencias de polipéptidos que comprenden una proteína/polipéptido efectora.

Los ácidos nucleicos o las proteínas de la descripción pueden construirse mediante un enfoque modular al ensamblar previamente unidades monoméricas y/o unidades de repetición en vectores objetivo que posteriormente pueden ensamblarse en un vector de destino final. En un aspecto, los polipéptidos de la descripción comprenden monómeros repetidos como se describe en la presente descripción y pueden construirse mediante un enfoque modular mediante el ensamblaje previo de unidades de repetición en vectores objetivo que posteriormente pueden ensamblarse en un vector de destino final. La descripción proporciona el polipéptido producido mediante este método, así como también las secuencias de ácido nucleico que codifican los polipéptidos y los organismos huésped y las células que comprenden dichas secuencias de ADN.

Las técnicas para modificar específicamente secuencias de ADN para obtener un codón específico para un aminoácido específico se conocen en la técnica. Los métodos para mutagénesis y alteraciones de polinucleótidos se han descrito ampliamente. Ver, por ejemplo, Kunkel (1985) Proc. Natl. Acad. Sci. EE.UU. 82:488-492; Kunkel y otros (1987) Methods in Enzymol. 154:367-382; Patente de Estados Unidos Núm. 4,873,192; Walker y Gaastra, eds. (1983) Techniques in Molecular Biology (MacMillan Publishing Company, Nueva York) y las referencias citadas en ellas.

Los siguientes ejemplos proporcionan métodos para construir nuevas unidades de repetición y probar las actividades de unión específica de unidades de repetición construidas artificialmente que reconocen específicamente pares de bases en una secuencia de ADN objetivo. El número de unidades de repetición que se usarán en un dominio de repetición puede determinarlo un experto en la técnica mediante experimentación de rutina. Generalmente, al menos 1,5 unidades de repetición se consideran como un mínimo, aunque típicamente se usarán al menos aproximadamente 8 unidades de repetición. Las unidades de repetición no tienen que ser unidades de repetición completas, ya que pueden usarse unidades de repetición de la mitad del tamaño. Además, los métodos y polipéptidos descritos en la presente descripción dependen de dominios de repetición con un número particular de unidades de repetición. De esta forma, un polipéptido de la descripción puede comprender, por ejemplo, 1,5, 2, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5, 10, 10,5, 11, 11,5, 12, 12,5, 13, 13,5, 14, 14,5, 15, 15,5, 16, 16,5, 17, 17,5, 18, 18,5, 19, 19,5, 20, 20,5, 21, 21,5, 22, 22,5, 23, 23,5, 24, 24,5, 25, 25,5, 26, 26,5, 27, 27,5, 28, 28,5, 29, 29,5, 30, 30,5, 31, 31,5, 32, 32,5, 33, 33,5, 34, 34,5, 35, 35,5, 36, 36,5, 37, 37,5, 38, 38,5, 39, 39,5, 40, 40,5, 41, 41,5, 42, 42,5, 43, 43,5, 44, 44,5, 46, 46,5, 47, 47,5, 48, 48,5, 49, 49,5, 50, 50,5 o más unidades de repetición.

En la descripción, pueden diseñarse polipéptidos que comprenden un dominio de repetición con unidades de repetición en donde en las unidades de repetición se incluyen regiones hipervariables que determinan el reconocimiento de un par de bases en una secuencia de ADN objetivo. En una modalidad, cada unidad de repetición incluye una región hipervariable que determina el reconocimiento de un par de bases en una secuencia de ADN objetivo. En una modalidad adicional, se

incluyen 1 o 2 unidades de repetición en un dominio de repetición que no reconocen específicamente un par de bases en una secuencia de ADN objetivo. Si se tiene en cuenta el código de reconocimiento encontrado por los inventores, es factible una disposición modular de unidades de repetición en donde cada unidad de repetición es responsable del reconocimiento específico de un par de bases en una secuencia de ADN objetivo. En consecuencia, una secuencia de unidades de repetición corresponde a una secuencia de pares de bases en una secuencia de ADN objetivo de modo que 1 unidad de repetición coincide con un par de bases.

La descripción proporciona un método para reconocer selectivamente un par de bases en una secuencia de ADN objetivo mediante un polipéptido, en donde dicho polipéptido comprende al menos un dominio de repetición que comprende unidades de repetición, en donde en dichas unidades de repetición cada una comprende al menos una región de RVD que determina el reconocimiento de un par de bases o nucleótido en dicha secuencia de ADN objetivo. Más específicamente, los inventores han determinado los aminoácidos en un polipéptido de unión a ADN responsables del reconocimiento selectivo de pares de bases en una secuencia de ADN objetivo. Con la dilucidación del código de reconocimiento, se ha determinado un principio general para reconocer pares de bases específicos en una secuencia de ADN objetivo mediante aminoácidos seleccionados en un polipéptido. Los inventores han descubierto que distintos tipos de monómeros que forman parte de un conjunto de unidades de repetición (o polímero) de longitud variable de aminoácidos tienen la capacidad de reconocer un par de bases definido/específico. Dentro de cada unidad de repetición que forma un dominio de repetición, una región de RVD es responsable del reconocimiento específico de un par de bases en una secuencia de ADN objetivo.

Por lo tanto, la descripción proporciona no solo un método para reconocer selectivamente un par de bases en una secuencia de ADN objetivo mediante un polipéptido que comprende al menos un dominio de repetición que comprende unidades de repetición, sino también métodos en donde pueden generarse secuencias de ADN objetivo que se reconocen selectivamente mediante dominios de repetición en un polipéptido. Estos polipéptidos son útiles como herramientas de biología molecular para clonar, mutagenizar o de cualquier otra manera alterar una secuencia de ácido nucleico aislada u otra secuencia *in vivo* en un laboratorio. Esto proporciona un medio eficiente de mutagénesis selectiva.

La descripción proporciona, además, un método para construir y/o crear polipéptidos que reconocen secuencias de ADN específicas. Estos polipéptidos de la descripción comprenden monómeros repetidos y pueden construirse mediante un enfoque modular mediante el ensamblaje previo de unidades de repetición en vectores objetivo que posteriormente pueden ensamblarse en un vector de destino final. En algunas modalidades, los constructos de ADN se optimizan en codones para producir y/o secretar de forma recombinante los polipéptidos descritos en la presente descripción. Puede usarse cualquier sistema recombinante en la técnica para producir la proteína recombinante. Los ejemplos incluyen células de baculovirus, otras células eucariotas tales como células de mamífero o células bacterianas.

Siempre y cuando se conozca una secuencia de ADN objetivo y para la cual se desee el reconocimiento por parte de una proteína, el experto en la técnica puede construir específicamente una serie modular de unidades de repetición, lo que incluye secuencias de aminoácidos de reconocimiento específicas, y ensamblar estas unidades de repetición en un polipéptido en el orden apropiado para permitir el reconocimiento y la unión a la secuencia de ADN objetivo deseada. Cualquier polipéptido puede modificarse mediante la combinación con un dominio de unión al ADN de unidades de repetición modulares descrito en la presente descripción. Dichos ejemplos incluyen polipéptidos que son proteínas activadoras y represoras de la transcripción, proteínas mediadoras de resistencia, nucleasas, topoisomerasas, ligasas, integrasas, recombinasas, resolvasas, metilasas, acetilasas, desmetilasas, desacetilasas y cualquier otro polipéptido capaz de modificar ADN, ARN o proteínas.

El dominio de unión al ADN de la unidad de repetición modular de la descripción puede combinarse con señales de localización del compartimento celular tales como señales de localización nuclear, para funcionar en cualquier otra región reguladora, lo que incluye, entre otras, regiones reguladoras transcripcionales y regiones de terminación traduccional.

En una modalidad adicional de la descripción, estas unidades de repetición diseñadas modularmente se combinan con un dominio de endonucleasa capaz de escindir el ADN cuando se acerca al ADN como resultado de la unión por el dominio de repetición. Se conoce que dichas rupturas endonucleolíticas estimulan la tasa de recombinación homóloga en eucariotas, lo que incluye hongos, plantas y animales. La capacidad de estimular la recombinación homóloga en un sitio específico como resultado de una ruptura endonucleolítica específica de sitio permite la recuperación de células transformadas que han integrado una secuencia de ADN de interés en el sitio específico, a una frecuencia mucho más alta de lo que es posible sin haber hecho la ruptura específica de sitio. Además, las rupturas endonucleolíticas tales como las provocadas por polipéptidos formados a partir de un dominio de repetición y un dominio de endonucleasa a veces se reparan por la maquinaria metabólica del ADN celular de una manera que altera la secuencia en el sitio de la ruptura, por ejemplo, al provocar una inserción corta o delección en el sitio de la ruptura en comparación con la secuencia inalterada. Estas alteraciones de la secuencia pueden provocar la inactivación de la función de un gen o proteína, por ejemplo, al alterar una secuencia codificante de proteína para hacer una proteína no funcional, modificar un sitio de empalme para que un transcripto génico no se escinda adecuadamente, crear un transcripto no funcional, cambiar la secuencia promotora de un gen para que ya no pueda transcribirse adecuadamente, etcétera.

Romper el ADN mediante el uso de endonucleasas específicas de sitio puede aumentar la tasa de recombinación homóloga en la región de la ruptura. En algunas modalidades, la endonucleasa Fok I (*Flavobacterium okeanoikotes*)

puede usarse en un efector para inducir rupturas de ADN. El dominio de la endonucleasa Fok I funciona independientemente del dominio de unión al ADN y corta un ADN bicatenario típicamente como un dímero (Li y otros (1992) Proc. Natl. Acad. Sci. EE.UU. 89 (10):4275-4279 y Kim y otros (1996) Proc. Natl. Acad. Sci. EE.UU. 93 (3):1156-1160). Además, se ha desarrollado un dímero FokI de cadena sencilla y también podría usarse (Mino y otros (2009) J. Biotechnol. 140:156-161). Podría construirse un efector que contenga un dominio de repetición para el reconocimiento de una secuencia de ADN objetivo deseada, así como también un dominio de endonucleasa FokI para inducir la ruptura del ADN en o cerca de la secuencia de ADN objetivo similar al trabajo previo realizado mediante el empleo de nucleasas de dedos de zinc (Townsend y otros (2009) Nature 459:442-445; Shukla y otros (2009) Nature 459, 437-441). El uso de dichos efectores podría permitir la generación de cambios dirigidos en los genomas que incluyen adiciones, deleciones y otras modificaciones, análogas a los usos reportados para las nucleasas de dedos de zinc según Bibikova y otros (2003) Science 300, 764; Urmov y otros (2005) Nature 435, 646; Wright y otros (2005) The Plant Journal 44:693-705; y las Patentes de Estados Unidos Núms. 7,163,824 y 7,001,768.

El dominio de la endonucleasa FokI puede clonarse por PCR a partir del ADN genómico de las bacterias marinas *Flavobacterium okeanoikoites* (ATCC) preparadas por métodos estándar. La secuencia de la endonucleasa FokI está disponible en Pubmed (Ace. Núm. M28828 y Acc. Núm. J04623). La endonucleasa I-Sce I de la levadura *Saccharomyces cerevisiae* se ha usado para producir rupturas de ADN que aumentan la tasa de recombinación homóloga. I-Sce I es una endonucleasa codificada por un intrón mitocondrial que tiene una secuencia de reconocimiento de 18 pb y, por lo tanto, una frecuencia muy baja de sitios de reconocimiento dentro de un ADN dado, incluso dentro de genomas grandes (Thierry y otros (1991) Nucleic Acids Res. 19 (1):189-190). La poca frecuencia de los sitios de escisión reconocidos por I-SceI la hace adecuada para potenciar la recombinación homóloga. Puede encontrarse una descripción adicional sobre el uso de I-Sce I para inducir dichas rupturas de ADN en la Solicitud de Patente de Estados Unidos 20090305402.

El sitio de reconocimiento para I-Sce I se ha introducido en una gama de sistemas diferentes. El posterior corte de este sitio con I-Sce I aumenta la recombinación homóloga en la posición donde se ha introducido el sitio. Se han obtenido frecuencias mejoradas de recombinación homóloga con sitios I-Sce I introducidos en el ADN extracromosómico en ovocitos de *Xenopus*, el genoma del ratón y el ADN genómico de la planta de tabaco *Nicotiana glauca*. Ver, por ejemplo, Segal y otros (1995) Proc. Natl. Acad. Sci. EE.UU. 92 (3):806-810; Choulila y otros (1995) Mol. Cell Biol. 15 (4):1968-1973; y Puchta y otros (1993) Nucleic Acids Res. 21 (22):5034-5040. Se apreciará que cualquier otro dominio de endonucleasa que funcione con dominios de unión a ADN heterólogos puede usarse en un efector y que la endonucleasa I-Sce I es uno de esos ejemplos no limitantes. La limitación del uso de endonucleasas que tienen un dominio de reconocimiento y unión a ADN tal como la I-Sce I es que el sitio de reconocimiento tiene que introducirse por métodos estándar de recombinación homóloga en la ubicación deseada antes del uso de dicha endonucleasa para mejorar la recombinación homóloga en ese sitio, si dicho sitio aún no está presente en la ubicación deseada. Se han informado métodos que permiten el diseño y la síntesis de nuevas endonucleasas, tales como modificar endonucleasas conocidas o crear versiones quiméricas de una o más de dichas endonucleasas, que reconocen nuevas secuencias de ADN objetivo, lo que allana de esta forma el camino para la generación de dichos dominios de endonucleasa modificados genéticamente para escindir secuencias de ADN objetivo endógenas de interés (Chevalier y otros (2002) Molecular Cell 10:895-905; documento WO2007/060495; documento WO2009/095793; Fajardo-Sánchez y otros (2008) Nucleic Acids Res. 36:2163-2173). Como tal, podría preverse que dichos dominios de endonucleasa podrían modificarse genéticamente de manera similar para hacer que la actividad de unión al ADN no sea funcional, pero dejar activa la función de escisión del ADN y usar dicho dominio de escisión de endonucleasa modificado genéticamente de manera similar en un efector para inducir rupturas de ADN similar al uso de FokI anteriormente. En dichas aplicaciones, el dominio de repetición del efector proporcionaría, preferentemente, el reconocimiento de la secuencia de ADN objetivo, pero la escisión del ADN se lograría mediante el dominio de endonucleasa modificado genéticamente.

Como se mencionó anteriormente, un efector incluye un dominio de repetición con reconocimiento específico para una secuencia objetivo específica deseada. En modalidades preferidas, el efector se une específicamente a una secuencia de ADN cromosómica endógena. La secuencia específica de ácido nucleico o, con mayor preferencia, la secuencia cromosómica endógena específica puede ser cualquier secuencia en una región de ácido nucleico donde se desea potenciar la recombinación homóloga. Por ejemplo, la región de ácido nucleico puede ser una región que contiene un gen en el que se desea introducir una mutación, tal como una mutación o delección puntual, o una región en la que se desea introducir un gen que confiere un fenotipo deseado.

Otras modalidades se refieren a métodos para generar una planta modificada en la que se ha introducido una adición deseada. Los métodos pueden incluir obtener una célula vegetal que incluye una secuencia de ADN objetivo endógena en la que se desea introducir una modificación; generar un corte bicatenario dentro de la secuencia de ADN objetivo endógena con un efector que incluye un dominio de repetición que se une a una secuencia de ADN objetivo endógeno y un dominio de endonucleasa; introducir un ácido nucleico exógeno que incluye una secuencia homóloga a al menos una porción del ADN objetivo endógeno en la célula vegetal en condiciones que permitan que se produzca una recombinación homóloga entre el ácido nucleico exógeno y la secuencia de ADN objetivo endógeno; y generar una planta a partir de la célula vegetal en la que se ha producido una recombinación homóloga. Otras modalidades se refieren a células y plantas modificadas genéticamente creadas de acuerdo con el método descrito anteriormente y en la presente descripción. Debe señalarse que la secuencia de ADN objetivo podría ser artificial o de origen natural. Se apreciará que dichos métodos podrían usarse en cualquier organismo no humano (dichos organismos no limitantes incluyen animales, hongos,

oomicetos, bacterias y virus) mediante el uso de técnicas y métodos conocidos en la técnica y usados para tales fines en dichos organismos.

En una modalidad adicional de la descripción, estos dominios de repetición diseñados modularmente se combinan con uno o más dominios responsables de la modulación o control de la expresión de un gen, por ejemplo, de genes vegetales, genes animales, genes fúngicos, genes de oomicetos, genes virales o genes humanos. Los métodos para modular la expresión génica mediante la generación de polipéptidos de unión a ADN que contienen dominios de dedos de zinc se conocen en la técnica (Patentes de Estados Unidos Núms. 7,285,416, 7,521,241, 7,361,635, 7,273,923, 7,262,054, 7,220,719, 7,070,934, 7,013,219, 6,979,539, 6,933,113, 6,824,978). Por ejemplo, estos efectores de la familia de similares a *Ralstonia* se modifican para unirse a secuencias de ADN objetivo específicas. Dichos polipéptidos podrían ser, por ejemplo, activadores de la transcripción o proteínas represoras de la transcripción que se modifican mediante el método de la descripción para unirse específicamente a las regiones de control genético en un promotor u otra región reguladora de un gen de interés para activar, reprimir o de cualquier otra manera modular la transcripción de dicho gen.

En aún otra modalidad de la descripción, las secuencias de ADN objetivo se modifican para reconocerse específicamente por un dominio de repetición de origen natural o por un dominio de repetición modificado. Como un ejemplo, las secuencias de ADN objetivo para miembros de la familia similar a *Ralstonia* pueden insertarse en promotores para generar nuevos promotores controlables que pueden inducirse mediante el efector correspondiente. Los sistemas inducibles secundarios pueden construirse mediante el uso de un activador de la transcripción y un gen objetivo, en donde el activador de la transcripción es un polipéptido en donde dicho polipéptido comprende al menos un dominio de repetición que comprende unidades de repetición de la descripción que se unen a dicho gen objetivo e inducen la expresión. El activador de la transcripción y el gen objetivo pueden introducirse en una línea celular pero también pueden presentarse en diferentes líneas celulares y después incorporarse por introgresión. En una modalidad adicional, las plantas resistentes a enfermedades pueden construirse mediante la inserción de la secuencia de ADN objetivo de un polipéptido que contiene un dominio de repetición de la descripción frente a un gen que después de la expresión conduce a una reacción de defensa de la planta al activar un gen mediador de resistencia.

En una modalidad adicional, los polipéptidos de unión a ADN personalizados pueden construirse el reordenar los tipos de unidades de repetición, lo que permite de esta manera la generación de dominios de repetición con una especificidad de unión al ADN objetivo novedosa. Las unidades de repetición individuales son casi idénticas a nivel de ADN, lo que impide las estrategias clásicas de clonación. Esto proporciona una estrategia rápida y económica para ensamblar polipéptidos personalizados con dominios de repetición. Para mejorar la versatilidad de clonación de dichos polipéptidos, se diseñó un método de ensamblaje de dos etapas. Este método se usó para ensamblar polipéptidos con nuevos tipos de repetición para estudiar su reconocimiento de ADN objetivo y especificidad de unión.

En resumen, cualquier secuencia de ADN puede modificarse para permitir la unión por un polipéptido que contiene un dominio de repetición mediante la introducción de pares de bases en cualquier región del ADN o regiones específicas de un gen o un elemento de control genético, para dirigirse específicamente a un polipéptido que tiene un dominio de repetición compuesto por unidades de repetición que se unirá a dicha secuencia de ADN modificada para facilitar el reconocimiento específico y la unión entre sí.

En algunas modalidades, los polipéptidos pueden fabricarse sintéticamente mediante el uso de químicas de aminoácidos conocidas familiares para un experto en síntesis de química orgánica. Dichos procedimientos incluyen tanto procedimientos en solución como en fase sólida, por ejemplo, mediante el uso de metodologías Boc y Fmoc. Los compuestos pueden sintetizarse mediante el uso de técnicas de síntesis en fase sólida. Los β -aminoácidos N-protectados con Fmoc pueden usarse para sintetizar péptidos poli- α/β mediante procedimientos convencionales de síntesis en fase sólida manuales en condiciones estándar en cualquier número de soportes sólidos, lo que incluye la resina de cloruro de ortoclorotritilo. La esterificación de Fmoc- β -aminoácidos con la resina de ortoclorotritilo puede realizarse de acuerdo con el método de Barlos y otros, Tetrahedron Lett., 1989, 30, 3943. La resina (150 mg, 1,05 mmol Cl) se hincha en 2 ml de CH_2Cl_2 durante 10 min. Una solución del β -aminoácido protegido con Fmoc en CH_2Cl_2 e $i\text{Pr}_2\text{EtN}$ se añaden después sucesivamente y la suspensión se mezcla en atmósfera de argón durante 4 h. Posteriormente, la resina se filtra y se lava con $\text{CH}_2\text{Cl}_2/\text{MeOH}/i\text{Pr}_2\text{EtN}$ (17:2:1, 3x3 min), CH_2Cl_2 (3x3 min), DMF (2x3 min), CH_2Cl_2 (3x3 min) y MeOH (2x3 min). La sustitución de la resina se determina en una muestra de 3 mg al medir la absorbancia del aducto de dibenzofulveno a 300 nm. El grupo Fmoc se elimina mediante el uso de piperidina al 20 % en DMF (4 ml, 2x20 min) con burbujeo de Ar. A continuación, la resina se filtra y se lava con DMF (6x3 min). Para cada etapa de acoplamiento, se añaden sucesivamente una solución del β -aminoácido (3 equiv.), BOP (3 equiv.) y HOBt (3 equiv.) en DMF (2 ml) e $i\text{Pr}_2\text{EtN}$ (9 eq) se añaden sucesivamente a la resina y la suspensión se mezcla durante 1 h en atmósfera de Ar. La monitorización de la reacción de acoplamiento se realiza con ácido 2,4,6-trinitrobenzenosulfónico (TNBS) (W.S. Hancock y J.E. Battersby, Anal. Biochem. (1976), 71, 260). En el caso de una prueba TNBS positiva (lo que indica un acoplamiento incompleto), la suspensión se deja reaccionar durante 1 h adicional. A continuación, la resina se filtra y se lava con DMF (3x3 min) antes de la siguiente etapa de desprotección de Fmoc. Después de la eliminación del último grupo protector Fmoc, la resina se lava con DMF (6x3 min), CH_2Cl_2 (3x3 min), Et_2O (3x3 min) y se seca al vacío durante 3 h. Finalmente, los péptidos se separan de la resina mediante el uso de TFA al 2 % en CH_2Cl_2 (2 ml, 5x15 min) en atmósfera de Ar. El solvente se elimina y los residuos oleosos se trituran en éter para obtener los polipéptidos crudos. Los compuestos se purifican posteriormente por HPLC.

- La descripción proporciona, además, un método para la modulación dirigida de la expresión génica mediante la construcción de unidades de repetición modulares específicas para una secuencia de ADN objetivo de interés, modificar un polipéptido mediante la adición de dichos monómeros repetidos para permitir que dicho polipéptido reconozca ahora el ADN objetivo, introducir o expresar dicho polipéptido modificado en una célula procariota o eucariota para permitir que dicho polipéptido modificado reconozca la secuencia de ADN objetivo, y modular la expresión de dicho gen objetivo en dicha célula como resultado de dicho reconocimiento.
- La descripción proporciona, además, un método para la modificación dirigida de una secuencia de ADN objetivo mediante la construcción de un polipéptido que incluye al menos un dominio de repetición de la descripción que reconoce dicha secuencia de ADN objetivo y que dicho polipéptido contiene, además, un dominio funcional capaz de modificar el ADN objetivo (tal como a través de la recombinación específica de sitio, la restricción o la integración de las secuencias objetivo del donante) lo que permite de esta manera modificaciones del ADN dirigidas en genomas complejos.
- La descripción proporciona, además, la producción de polipéptidos modificados que incluyen al menos un dominio de repetición que comprende unidades de repetición, en donde una región hipervariable dentro de cada una de las unidades de repetición determina el reconocimiento selectivo de un par de bases en una secuencia de ADN objetivo. En una modalidad adicional de la descripción, se proporciona ADN que codifica un polipéptido que contiene un dominio de repetición como se describió anteriormente.
- En otra modalidad de la descripción, se proporciona ADN que se modifica para incluir uno o más pares de bases ubicadas en una secuencia de ADN objetivo de modo que cada uno de dichos pares de bases pueda reconocerse específicamente por un polipéptido que incluye un dominio de repetición que tiene unidades de repetición correspondientes, cada unidad de repetición comprende una región hipervariable que determina el reconocimiento del par de bases correspondiente en dicho ADN.
- En aún otra modalidad de la descripción, se proporcionan usos de esos polipéptidos y ADN. Adicionalmente se proporcionan plantas, partes de plantas, semillas, células vegetales y otras células huésped no humanas transformadas con las moléculas de ácido nucleico aisladas de la descripción y las proteínas o polipéptidos codificados por las secuencias codificantes de la descripción. Aún más, los polipéptidos y el ADN descritos en la presente descripción pueden introducirse en células animales y humanas, así como también en células de otros organismos como hongos o plantas.
- En resumen, la descripción se centra en un método para reconocer selectivamente pares de bases en una secuencia de ADN objetivo mediante un polipéptido, en donde dicho polipéptido comprende al menos un dominio de repetición que comprende unidades de repetición, en donde cada unidad de repetición contiene una región hipervariable que determina el reconocimiento de un par de bases en dicha secuencia de ADN objetivo, en donde las unidades de repetición consecutivas corresponden a pares de bases consecutivas en dicha secuencia de ADN objetivo.
- En algunas modalidades, la descripción se refiere a una célula humana que comprende una cualquiera o una combinación de proteínas o secuencias de ácido nucleico descritas en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula humana que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula animal no humana que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula de insecto que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula vegetal que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula de pez que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula de mamífero que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción. En algunas modalidades, la descripción se refiere a células tales como una célula eucariota que comprende una mutación, un gen heterólogo, una variante u otra modificación genética provocada por la introducción de uno o más ácidos nucleicos o polipéptidos descritos en la presente descripción.
- En otro aspecto, se proporciona un método para modular la expresión de un gen objetivo en una célula. La célula puede ser, preferentemente, una célula vegetal, una célula humana, una célula animal, una célula fúngica o cualquier otra célula viva. Las células contienen un polipéptido en donde dicho polipéptido comprende al menos un dominio de repetición que comprende unidades de repetición, y estas unidades de repetición contienen una región hipervariable y cada unidad de repetición es responsable del reconocimiento de 1 par de bases en dicha secuencia de ADN objetivo. Dicho polipéptido se introduce ya sea como ADN que codifica para el polipéptido o el polipéptido se introduce per se en la célula mediante métodos conocidos en la técnica. Independientemente de cómo se haya introducido, el polipéptido debe incluir al menos un dominio de repetición que reconozca específicamente y se una, preferentemente, a una secuencia de ADN objetivo de

pares de bases y module la expresión de un gen objetivo. En una modalidad preferida, todas las unidades de repetición contienen una región hipervariable que determina el reconocimiento de pares de bases en una secuencia de ADN objetivo.

Los ejemplos de secuencias peptídicas que pueden unirse a un polipéptido o RTN de la descripción, para facilitar la captación de efectores en las células, incluyen, pero no se limitan a: un péptido de 11 aminoácidos de la proteína tat del VIH; una secuencia peptídica de 20 residuos que corresponde a los 84 103 aminoácidos de la proteína p16 (ver Fahraeus y otros (1996) *Current Biology* 6:84); la tercera hélice del homeodominio de 60 aminoácidos de longitud de Antennapedia (Derossi y otros (1994) *J. Biol. Chem.* 269:10444); la región h de un péptido señal tal como la región h del factor de crecimiento de fibroblastos de Kaposi (K-FGF); o el dominio de translocación VP22 de HSV (Elliot y O'Hare (1997) *Cell* 88:223 233). Otros restos químicos adecuados que proporcionan una captación celular mejorada también pueden unirse químicamente a efectores. Como se describe en la presente descripción, los efectores pueden diseñarse para reconocer cualquier sitio objetivo adecuado, para la regulación de la expresión de cualquier gen endógeno de elección. Los ejemplos de genes endógenos adecuados para la regulación incluyen VEGF, CCR5, ER.alfa., Her2/Neu, Tat, Rev, HBV C, S, X y P, LDL-R, PEPCK, CYP7, fibrinógeno, ApoB, Apo E, Apo(a), renina, NF-kappa.B, I-kappa.B, TNF-alfa, ligando FAS, proteína precursora amiloide, factor natural auricular, ob-leptina, ucp-1, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-12, G-CSF, GM-CSF, Epo, PDGF, PAF, p53, Rb, hemoglobina fetal, distrofina, eutrofina, GDNF, NGF, IGF-1, receptores VEGF flt y flk, topoisomerasa, telomerasa, bcl-2, ciclinas, angiostatina, IGF, ICAM-1, STATS, c-myc, c-myb, TH, PTI-1, poligalacturonasa, EPSP sintasa, FAD2-1, delta-12 desaturasa, delta-9 desaturasa, delta-15 desaturasa, acetil-CoA carboxilasa, acil-ACP-tioesterasa, ADP-glucosa pirofosforilasa, almidón sintasa, celulosa sintasa, sacarosa sintasa, genes asociados a la senescencia, quelantes de metales pesados, hidroperóxido liasa de ácidos grasos, genes virales, genes protozoarios, genes fúngicos y genes bacterianos. En general, los genes adecuados para regularse incluyen citocinas, linfocinas, factores de crecimiento, factores mitogénicos, factores quimiotácticos, factores ontoactivos, receptores, canales de potasio, proteínas G, moléculas de transducción de señales, genes de resistencia a enfermedades y otros genes relacionados con enfermedades.

Las moléculas de toxina también tienen la capacidad de transportar polipéptidos a través de las membranas celulares. A menudo, dichas moléculas se componen de al menos dos partes (llamadas "toxinas binarias"): un dominio o polipéptido de translocación o unión y un dominio o polipéptido de toxina separado. Típicamente, el dominio o polipéptido de translocación se une a un receptor celular, y después la toxina se transporta a la célula. Se han usado varias toxinas bacterianas, lo que incluye la toxina iota *Clostridium perfringens*, la toxina diftérica (DT), la exotoxina A de *Pseudomonas* (PE), la toxina pertussis (PT), la toxina de *Bacillus anthracis* y la adenilato ciclasa pertussis (CYA), en intentos de suministrar péptidos al citosol celular como fusiones internas o amino terminales (Arora y otros (1993) *J. Biol. Chem.* 268:3334 3341; Perelle y otros (1993) *Infect. Immun.* 61:5147 5156 (1993); Stenmark y otros (1991) *J. Cell Biol.* 113:1025 1032 (1991); Donnelly y otros (1993) *Proc. Natl. Acad. Sci. EE.UU.* 90:3530 3534; Carbonetti y otros (1995) *Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295; Sebo y otros (1995) *Infect. Immun.* 63:3851 3857; Klimpel y otros (1992) *Proc. Natl. Acad. Sci. EE.UU.* 89:10277 10281; y Novak y otros (1992) *J. Biol. Chem.* 267:17186 17193).

Los efectores pueden introducirse, además, en una célula animal, preferentemente, una célula de mamífero, mediante liposomas y derivados de liposomas tales como inmunoliposomas. El término "liposoma" se refiere a vesículas compuestas de una o más bicapas lipídicas ordenadas concéntricamente, que encapsulan una fase acuosa. La fase acuosa contiene típicamente el compuesto que se suministrará a la célula, en este caso un efector. El liposoma se fusiona con la membrana plasmática, lo que libera de esta manera el efector en el citosol. Alternativamente, el liposoma se fagocita o capta por la célula en una vesícula de transporte. Una vez en el endosoma o fagosoma, el liposoma se degrada o se fusiona con la membrana de la vesícula de transporte y libera su contenido.

La descripción se refiere particularmente al campo de la tecnología vegetal y agrícola. En un aspecto, la descripción se dirige a un método para modular la expresión de un gen objetivo en células vegetales, método que comprende proporcionar a las plantas vegetales un polipéptido modificado de acuerdo con la descripción, dicho polipéptido es capaz de reconocer específicamente una secuencia de nucleótidos objetivo, o una cadena complementaria de esta, dentro de un gen objetivo, y que permite que dicho polipéptido reconozca y se una particularmente a dicha secuencia de nucleótidos objetivo, de manera que se modula la expresión de dicho gen objetivo en dichas células vegetales.

El polipéptido puede proporcionarse a las células vegetales mediante cualquier método adecuado conocido en la técnica. Por ejemplo, la proteína puede añadirse de manera exógena a las células vegetales y las células vegetales se mantienen en condiciones tales que el polipéptido se introduce en la célula vegetal, se une a la secuencia de nucleótidos objetivo y regula la expresión del gen objetivo en las células vegetales. Alternativamente, una secuencia de nucleótidos, por ejemplo, ADN o ARN, que codifica el polipéptido, puede expresarse en las células vegetales y las células vegetales se mantienen en condiciones tales que el polipéptido expresado se une a la secuencia de nucleótidos objetivo y regula la expresión del gen objetivo en las células vegetales.

Un método preferido para modular la expresión de un gen objetivo en células vegetales comprende las siguientes etapas: a) proporcionar a las plantas vegetales un sistema de expresión para un polipéptido modificado de acuerdo con la descripción, dicho polipéptido es capaz de reconocer específicamente y, preferentemente, unirse a una secuencia de nucleótidos objetivo, o una cadena complementaria de esta, dentro de un elemento de control de la expresión de un gen objetivo, preferentemente, un promotor; y b) cultivar dichas células vegetales en condiciones en donde dicho polipéptido

se produce y se une a dicha secuencia de nucleótidos objetivo, de manera que se modula la expresión de dicho gen objetivo en dichas células vegetales.

Cualquier secuencia de nucleótidos objetivo puede modularse por el presente método. Por ejemplo, la secuencia de nucleótidos objetivo puede ser endógena o exógena al gen objetivo. En una modalidad de la descripción, la secuencia de nucleótidos objetivo puede presentarse en una célula viva o presentarse *in vitro*. En una modalidad específica, la secuencia de nucleótidos objetivo es endógena a la planta. La secuencia de nucleótidos objetivo puede ubicarse en cualquier lugar adecuado en relación con el gen objetivo. Por ejemplo, la secuencia de nucleótidos objetivo puede ubicarse aguas arriba o aguas abajo de la región codificante del gen objetivo. Alternativamente, la secuencia de nucleótidos objetivo se ubica dentro de la región de codificación del gen objetivo. Preferentemente, la secuencia de nucleótidos objetivo es un promotor de un gen.

Cualquier gen objetivo puede modularse mediante el presente método. Por ejemplo, el gen objetivo puede codificar un producto que afecta la biosíntesis, modificación, tráfico celular, metabolismo y degradación de un péptido, una proteína, un oligonucleótido, un ácido nucleico, una vitamina, un oligosacárido, un carbohidrato, un lípido o una pequeña molécula. Además, los efectores pueden usarse para modificar genéticamente plantas para rasgos tales como aumento de la resistencia a enfermedades, modificación de polisacáridos estructurales y de almacenamiento, sabores, proteínas y ácidos grasos, maduración de frutos, rendimiento, color, características nutricionales, capacidad de almacenamiento mejorada y similares.

Por lo tanto, la descripción proporciona un método para alterar la expresión de un gen de interés en una célula objetivo, que comprende: determinar (si es necesario) al menos parte de la secuencia de ADN de la región estructural y/o una región reguladora del gen de interés; diseñar un polipéptido que incluye las unidades de repetición modificadas como se describe en la presente descripción para reconocer pares de bases específicos en el ADN de secuencia conocida, y provocar que dicho polipéptido modificado se presente en la célula objetivo (preferentemente, en el núcleo de esta). (Será evidente que la secuencia de ADN no necesita determinarse si ya se conoce).

La descripción proporciona, además, kits que comprenden: (1) cualquiera de los vectores mencionados anteriormente o (2) cualquiera de las proteínas o polipéptidos mencionados anteriormente. La descripción proporciona, además, kits que comprenden: (1) cualquiera de los vectores mencionados anteriormente o (2) cualquiera de las proteínas o polipéptidos mencionados anteriormente; y (3) cualquiera de las células mencionadas anteriormente (ya sea modificadas o no modificadas) descritas en la presente descripción.

En otras modalidades, la descripción se refiere a kits que se usan para producir mutaciones específicas de sitio en células madre, que pueden usarse para generar organismos modificados genéticamente. Los kits incluyen típicamente una o más tecnologías de ingeniería genética específicas de sitio, tales como los RTN. El kit puede contener, además, uno o más conjuntos de células madre o células embrionarias no humanas para la modificación específica de sitio. Las células madre pueden incluir, pero no se limitan a, células madre espermátogoniales (SSC), así como también los medios y las condiciones necesarias para el cultivo de las SSC. Los kits pueden incluir secuencias exógenas para la introducción genómica específica de sitio, tales como, pero sin limitarse a, genes indicadores o marcadores seleccionables. Los kits pueden incluir instrucciones para (i) introducir los RTN en las células madre (ii) identificar las células madre que se han modificado específicamente en el sitio (iii) cultivar las células madre modificadas específicamente en el sitio en los medios o condiciones necesarias y en las cantidades requeridas por las células madre para producir organismos modificados genéticamente (iv) usar las células madre cultivadas para producir un organismo modificado genéticamente (v) identificar qué organismos o progenie albergan la mutación específica de sitio de interés.

En alguna modalidad, la descripción proporciona un kit que incluye una población mixta de SSC modificadas genéticamente diferentes o distintas que pueden hacerse a la medida. La población mixta de SSC modificadas genéticamente puede proporcionarse en cantidades adecuadas para la inyección directa en un receptor masculino estéril para la producción de múltiples organismos modificados genéticamente en una sola etapa. La población mixta de SSC modificadas genéticamente separadas o distintas puede consistir en al menos dos SSC modificadas genéticamente, al menos dos SSC modificadas genéticamente, al menos tres SSC modificadas genéticamente, al menos cuatro SSC modificadas genéticamente, al menos cinco SSC modificadas genéticamente, al menos seis SSC modificadas genéticamente, al menos siete SSC modificadas genéticamente, al menos ocho SSC modificadas genéticamente, al menos nueve SSC modificadas genéticamente, al menos diez SSC modificadas genéticamente, al menos veinte SSC modificadas genéticamente, al menos treinta SSC modificadas genéticamente, al menos cuarenta SSC modificadas genéticamente, al menos cincuenta SSC modificadas genéticamente, al menos cien SSC modificadas genéticamente, al menos mil SSC modificadas genéticamente, al menos diez mil SSC modificadas genéticamente, al menos treinta mil SSC modificadas genéticamente o con SSC modificadas genéticamente que albergan una modificación genética dentro de cada gen en el genoma del organismo.

En algunas modalidades, la descripción proporciona un kit que incluye una población mixta de células madre modificadas genéticamente diferentes o distintas o células embrionarias no humanas que pueden hacerse a la medida mediante cualquiera de los métodos descritos en la presente descripción. La población mixta de células madre modificadas genéticamente o células embrionarias no humanas puede proporcionarse en cantidades adecuadas para la inyección directa en un receptor masculino estéril para la producción de múltiples organismos modificados genéticamente en una

- sola etapa. La población mixta de células madre modificadas genéticamente separadas o distintas o células embrionarias no humanas puede consistir en al menos dos células madre modificadas genéticamente o células embrionarias no humanas, al menos dos células madre modificadas genéticamente o células embrionarias no humanas, al menos tres células madre modificadas genéticamente o células embrionarias no humanas, al menos cuatro células madre modificadas genéticamente o células embrionarias no humanas, al menos cinco células madre modificadas genéticamente o células embrionarias no humanas, al menos seis células madre modificadas genéticamente o células embrionarias no humanas, al menos siete células madre modificadas genéticamente o células embrionarias no humanas, al menos ocho células madre modificadas genéticamente o células embrionarias no humanas, al menos nueve células madre modificadas genéticamente o células embrionarias no humanas, al menos diez células madre modificadas genéticamente o células embrionarias no humanas, al menos veinte células madre modificadas genéticamente o células embrionarias no humanas, al menos treinta células madre modificadas genéticamente o células embrionarias no humanas, al menos cuarenta células madre modificadas genéticamente o células embrionarias no humanas, al menos cincuenta células madre modificadas genéticamente o células embrionarias no humanas, al menos cien células madre modificadas genéticamente o células embrionarias no humanas, al menos mil células madre modificadas genéticamente o células embrionarias no humanas, al menos diez mil células madre modificadas genéticamente o células embrionarias no humanas, al menos treinta mil células madre modificadas genéticamente o células embrionarias no humanas o con células madre modificadas genéticamente o células embrionarias no humanas que albergan una modificación genética dentro de cada gen en el genoma del organismo.
- En algunas modalidades, la descripción proporciona un kit que incluye uno o más conjuntos de células madre o células embrionarias no humanas o SSC para modificación específica de sitio. Los conjuntos de SSC pueden derivarse de organismos bien caracterizados que tienen diferentes estados de enfermedad. Las SSC pueden contener múltiples mutaciones, que pueden derivarse de modificaciones genéticas o de forma natural o por cualquier método. El kit puede incluir los medios y las condiciones para desarrollar las SSC en estado de enfermedad, así como también el macho receptor estéril para la producción de organismos modificados genéticamente.
- En algunas modalidades, la descripción proporciona un kit que incluye las herramientas necesarias para la derivación de líneas de SSC a partir de un organismo o muestra de tejido, así como también las herramientas necesarias para modificar genéticamente la SSC derivada y producir un organismo modificado genéticamente a partir de las SSC derivadas. El kit puede incluir herramientas de recolección de células tales como espermatoцитos para la cosecha, y herramientas de selección de SSC tal como la selección por laminina, y herramientas de propagación y crioconservación de SSC, así como también herramientas de validación de SSC que pueden incluir tinción de marcadores de superficie celular. El kit puede incluir, además, medios y condiciones para el crecimiento de las SSC, herramientas para la modificación genética de las SSC, así como también machos receptores estériles para la producción de organismos modificados genéticamente a partir de las SSC.
- En algunas modalidades, la descripción proporciona un kit que incluye SSC que se han generado a partir de células madre pluripotentes inducidas (iPS). Las células iPS pueden derivarse de diferentes fondos genéticos bien caracterizados, lo que incluye los estados de enfermedad, así como también los fondos genéticos regionales, de cepas y étnicos. El kit puede incluir, además, medios y condiciones para el crecimiento de las iPS, herramientas para la modificación genética de las iPS, así como también machos receptores estériles para la producción de organismos modificados genéticamente a partir de las iPS.
- Otros métodos para la transformación y generación de animales transgénicos o células modificadas aparecen en la Solicitud PCT Núm. de serie PCT/US2012/038465.
- Un aspecto de la descripción se refiere a un método para suministrar proteínas de fusión en una célula objetivo en donde la proteína de fusión comprende una proteína efectora. Crear una proteína de fusión implica aislar uno o más componentes polipeptídicos de los medios y posteriormente ligar el extremo terminal amino libre de un componente polipeptídico al extremo terminal carboxi de un segundo componente polipeptídico. En otras modalidades, las proteínas de fusión pueden prepararse mediante síntesis de polipéptidos simple y/o expresarse mediante la clonación de una secuencia de ácido nucleico en un vector de expresión. En el caso de la purificación de proteínas a partir de sistemas recombinantes basados en células que expresan constructos de expresión de la descripción, un experto en la técnica puede identificar señales de secreción compatibles que pueden determinarse fácilmente para cualquier sistema de secreción de tipo III particular que se empleará si dichos constructos de expresión se transforman en células huésped bacterianas para la producción de proteínas. Al identificar las proteínas que normalmente se secretan por el sistema de secreción de tipo III, es posible preparar mutantes de delección que carecen de diversos fragmentos de la proteína de longitud completa que normalmente se secreta mediante el sistema de secreción. Mediante el uso de anticuerpos marcados generados contra epítopos de los diversos fragmentos de delección que se expresan (es decir, epítopos N-terminales, epítopos C-terminales, etcétera), es posible identificar mutantes de delección que se secretan y aquellos que no se secretan. Por lo tanto, los dominios de proteínas necesarios para la secreción de la proteína de longitud completa pueden identificarse fácilmente. Una vez que los dominios de proteínas se han identificado y secuenciado, pueden usarse como señales de secreción en proteínas de fusión de la descripción.
- Típicamente, la señal de secreción es un dominio N-terminal de una proteína que se secreta normalmente por el sistema de secreción de tipo III particular, por ejemplo, una secuencia de 201 aminoácidos del dominio N-terminal de la proteína

DspE de *Erwinia amylovora* (ver, por ejemplo, la Solicitud de Patente de Estados Unidos Núm. de serie 09/350,852, presentada el 9 de julio de 1999). La señal de secreción de 201 aminoácidos de DspE de *Erwinia amylovora* es compatible con el sistema de secreción de hapiN de *Erwinia amylovora*. Otras señales de secreción que son compatibles con diversos sistemas de secreción de tipo III se han descrito en la técnica y otras se identifican continuamente.

La proteína efectora purificada puede obtenerse por varios métodos. La proteína o polipéptido se produce, preferentemente, en forma purificada (al menos aproximadamente 80 %, 90 %, pura) mediante técnicas convencionales. Debido a que las células huésped recombinantes expresan un sistema de secreción de tipo III, la proteína o polipéptido se secreta en el medio de crecimiento de las células huésped recombinantes. En dichos casos, para aislar la proteína, las células huésped recombinantes se propagan, el medio de crecimiento se centrifuga para separar los componentes celulares del sobrenadante que contiene la proteína o polipéptido secretado, y se extrae el sobrenadante. A continuación, el sobrenadante se somete a precipitación secuencial con sulfato de amonio. La fracción que contiene el polipéptido o proteína se somete a filtración en gel en una columna de dextrano o poli(acrilamida) del tamaño apropiado para separar las proteínas. Si es necesario, la fracción de proteína puede purificarse adicionalmente por HPLC.

Las proteínas efectoras que portan dominios de transducción de proteínas también pueden prepararse independientemente del sistema de secreción de tipo III mediante el uso de técnicas actuales del estado del arte para preparar grandes cantidades de proteínas purificadas a partir de células de *E. coli* recombinantes. Dichas técnicas emplean promotores fuertes e inducibles y etiquetas peptídicas, tales como His6, para la purificación por afinidad en una sola etapa de la proteína recombinante de los lisados de células de *E. coli*.

En una modalidad, la célula objetivo es una célula eucariota. Las células eucariotas incluyen aquellas del cultivo de tejidos, tales como las células HeLa, o en animales enteros, como las administradas al ratón mediante inyección intraperitoneal (Schwarze y otros, "Protein Transduction: Unrestricted Delivery into all Cells?" Trends Cell Biol. 10:290-295 (2000)).

En un aspecto de la descripción, los elementos de unión o reconocimiento de ADN de la descripción pueden fusionarse de forma modular para crear una cadena de aminoácidos que se unen a una secuencia de ADN objetivo de elección. En otro aspecto de la descripción, el uno o más elementos de reconocimiento o unión al ADN pueden unirse a una o más proteínas efectoras. La proteína efectora puede producirse mediante un patógeno vegetal bacteriano, un patógeno animal o una bacteria de la rizosfera, lo que incluye, pero no se limita a los enteropatógenos *Escherichia coli*, *Salmonella typhimurium*, *Shigella* spp., *Yersinia* spp., *Pseudomonas syringae*, *Xanthomonas campestris*, *Ralstonia solanacearum*, *Erwinia amylovora*, *Pseudomonas fluorescens* y *Pseudomonas aeruginosa*. Las proteínas efectoras adecuadas incluyen un inductor de respuesta hipersensible, una proteína de virulencia, una proteína externa de respuesta hipersensible y dependiente de patogenicidad, una proteína de virulencia y una proteína de patogenicidad. Los ejemplos de proteínas efectoras HopPsyA AAF71481 (*P. syringae*), HopPtoA AF232006 (*P. syringae*), Tir BAA96815 (*E. coli*), ExoS AAG07228 (*P. aeruginosa*), ExoT AAG03434 (*P. aeruginosa*), ExoY AAG05579 (*P. aeruginosa*), SopE AAC02071 (*S. typhimurium*), SopB AAF21057 (SigA) (*S. typhimurium*), SipA CAA63302 (*S. typhimurium*), SptP AAC44349 (*S. typhimurium*), IpaB A34965 (*Shigella* spp.), IpaA AAA26525 (*Shigella* spp.), IpaD SI 5579 (*Shigella* spp.), YopE SI 4242 (*Yersinia* spp.), YopH AAC69768 (*Yersinia* spp.), YpkA AAC69765 (*Yersinia* spp.), YopJ AAC69766 (*YopP*) (*Yersinia* spp.), AvrPto AAA25728 (*P. syringae*), AvrBs2 AAD 1 1434 (*X. campestris*), y AvrBs3 CAA34257 (*X. campestris*) (ver, por ejemplo, Galan y otros, "Type III Secretion Machines: Bacterial Devices for Protein Delivery into Host Cells", Science 284:1322-1328 (1999)). En una modalidad, la proteína efectora es heteróloga (es decir, normalmente no está presente) con respecto a la célula objetivo.

De acuerdo con los propósitos de esta descripción, tal como se incorpora y se describe ampliamente en la presente descripción, esta descripción se refiere a métodos para ingeniería genética específica de sitio mediante el uso de RTN de células madre y gametos no humanos, lo que incluye, pero no se limita a, células pluripotentes, células totipotentes no humanas, células madre somáticas, células madre espermatozonales no humanas (SSC), células madre embrionarias (ES), células madre pluripotentes inducidas (iPS), embriones no humanos, células germinales no humanas, células germinales primordiales no humanas (PGC), células del tubo de las plantas, células de polen y esporas. Los métodos para la modificación genética específica de sitio de células madre incluyen, pero no se limitan a, el uso de proteínas de unión y escisión de ADN específicas de sitio tales como los RTN.

La modificación genética específica de sitio de las células madre resulta en una función alterada de los genes o productos genéticos y organismos modificados genéticamente, y se producen modelos de cultivo de células o tejidos a partir de estas células madre modificadas genéticamente. Las células madre y organismos modificados incluyen células y organismos con genes anulados y con genes introducidos.

En otro aspecto, la descripción se refiere a organismos no humanos modificados genéticamente creados por modificación genética específica de sitio mediante el uso de RTN o que incluye, pero no se limita a, mamíferos no humanos, lo que incluye ratas, ratones, cerdos, conejos, cobayas, perros, primates no humanos, minicerdos, así como también plantas, lo que incluye, pero no se limita a maíz, soja, arroz, papa, trigo, tabaco, tomate y *Arabidopsis*, así como también los descendientes y ancestros de dichos organismos.

En otra modalidad, la descripción proporciona kits que se usan para producir mutaciones específicas de sitio en células madre, que pueden usarse para generar organismos modificados genéticamente. Los kits incluyen típicamente una o más

tecnologías de ingeniería genética específicas de sitio, tales como los RTN. El kit puede comprender, además, uno o más conjuntos de células madre para la modificación específica de sitio. En algunas modalidades, las células madre pueden incluir, pero no se limitan a, células madre espermatozonales no humanas (SSC), así como también los medios y las condiciones necesarias para el cultivo de las SSC. En algunas modalidades, el kit comprende secuencias exógenas para la introducción genómica específica de sitio, tales como, pero sin limitarse a, genes indicadores o marcadores seleccionables. En algunas modalidades, el kit comprende instrucciones para (i) introducir los RTN (o la secuencia de ácido nucleico que codifica el RTN) en las células madre (ii) identificar las células madre que se han modificado específicamente en sitio por XTN (iii) cultivar células madre modificadas específicas de sitio en medios o condiciones necesarias y en los números requeridos para que las células madre no humanas produzcan organismos no humanos modificados genéticamente o efectúen la transmisión de la línea germinal en un animal no humano; (iv) usar o trasplantar las células madre cultivadas para producir un organismo modificado genéticamente; y/o (v) identificar qué organismos o progenie comprenden la mutación específica de sitio de interés. En algunas modalidades, una composición comprende una o más células madre o uno o más embriones no humanos, la una o más células madre o uno o más embriones no humanos comprenden una o más de las siguientes mutaciones: (i) una mutación de delección; (ii) una mutación de anulación; y/o (iii) una adición de una secuencia de ácido nucleico heterólogo; la una o más mutaciones de (i), (ii) y/o (iii) son mutaciones específicas de sitio provocadas por un RTN.

En algunas modalidades, la secuencia de ácido nucleico heterólogo se elige entre un marcador seleccionable o un gen ortólogo. En algunas modalidades, la una o más células madre se elige entre una célula madre espermatozonal no humana (SSC), una célula madre embrionaria o una célula madre pluripotente inducida.

En algunas modalidades, la una o más células madre se derivan del linaje de la línea germinal de un animal no humano o planta. En algunas modalidades, la una o más células madre o el uno o más embriones no humanos comprenden, además, al menos una repetición en tándem invertida de un transposón o una variante de este.

En algunas modalidades, la una o más células madre es una célula madre somática. En algunas modalidades, un organismo comprende una o más células madre, la una o más células madre comprenden una o más de las siguientes mutaciones: (i) una mutación de delección; (ii) una mutación de anulación; y/o (iii) una adición de una secuencia de ácido nucleico heterólogo; la una o más mutaciones de (i), (ii) y/o (iii) son mutaciones específicas de sitio provocadas por un RTN. En algunas modalidades, la una o más células madre comprenden una SSC.

En la descripción, la proteína efectora se fusiona con al menos uno de los elementos de reconocimiento de ADN descritos en la presente descripción o derivados o análogos funcionales de estos para producir una proteína de fusión.

Un aspecto de la descripción se refiere a un método para suministrar proteínas efectoras a una célula objetivo. Este método implica introducir en la célula objetivo una proteína efectora fusionada a un polipéptido que incluye al menos un dominio de repetición o elemento de reconocimiento de ADN que reconoce dicho ADN objetivo o derivados o análogos funcionales de este. Otro aspecto de la descripción se refiere a un constructo de ADN que incluye una primera molécula de ADN que codifica una proteína efectora y una segunda molécula de ADN asociada operativamente con la primera molécula de ADN y que codifica un polipéptido que incluye al menos un dominio de repetición de la descripción que reconoce dicho ADN o derivados o análogos funcionales de este.

El método de la descripción permite el suministro eficiente de proteínas efectoras a las células, en particular a las células de mamíferos. Este método permite, además, el suministro de proteínas efectoras para su uso en aplicaciones farmacéuticas, insecticidas, fungicidas, herbicidas y otras. En particular, permitirá el suministro de proteínas efectoras a los pacientes en forma de terapia con proteínas. La terapia con proteínas de longitud completa activas biológicamente permitirá el acceso a la especificidad evolutiva incorporada de estas proteínas para sus objetivos, lo que evita potencialmente de esta manera los efectos inespecíficos que a veces se ven con las terapias con moléculas pequeñas. Además, cuando se usa junto con vectores virales específicos de tejido, este uso permite el suministro dirigido de proteínas efectoras a células particulares con el beneficio adicional de la redistribución secundaria de la proteína efectora posterior al direccionamiento inicial. Puede encontrarse un precedente para este enfoque en un experimento en donde el dominio de transducción de proteínas VP22 se fusionó con la proteína supresora de tumores p53 (Phelan y otros, "Intercellular Delivery of Functional p53 by the Herpesvirus Protein VP22", Nat. Biotechnol. 16:440-443 (1998)).

En algunas modalidades, la descripción se refiere a una composición que comprende una o más secuencias de ácido nucleico con una secuencia de ácido nucleico insertada. En algunas modalidades, el ácido nucleico insertado comprende al menos un gen transcripcionalmente activo, que es una secuencia codificante que puede expresarse en condiciones intracelulares, por ejemplo, una secuencia codificante en combinación con cualquier elemento regulador de expresión requerido que se requiere para la expresión en el entorno intracelular de la célula objetivo, cuyo genoma se modifica mediante la unión y la posterior acción mediante cualquiera de los polipéptidos descritos en la presente descripción. Los genes transcripcionalmente activos de los ácidos nucleicos pueden comprender un dominio de nucleótidos, es decir, un módulo de expresión que incluye una secuencia codificante de nucleótidos unidos operativamente con elemento(s) regulador(es) o de mediación transcripcional requeridos. Los elementos de mediación transcripcional requeridos que pueden presentarse en el módulo de expresión incluyen, pero no se limitan a, promotores, potenciadores, elementos de señal de terminación y poliadenilación, elementos de señal de corte y empalme y similares. En algunas modalidades, la

una o más células madre comprenden, además, al menos una repetición terminal invertida de un transposón o variante de este.

- 5 En algunas modalidades, el módulo de expresión incluye elementos reguladores de la transcripción que proporcionan la expresión del gen en un amplio rango de huéspedes. Se conoce una variedad de dichas combinaciones, donde los elementos reguladores de la transcripción específicos incluyen, pero no se limitan a: elementos SV40, elementos reguladores de la transcripción derivados de la LTR del virus del sarcoma de Rous, elementos reguladores de la transcripción derivados de la LTR del citomegalovirus humano (CMV), promotores de hsp70 y similares.
- 10 En algunas modalidades, al menos un gen de transcripción activo o módulo de expresión presente en el ácido nucleico insertado actúa como un marcador seleccionable. Se han empleado una variedad de genes diferentes como marcadores seleccionables, y el gen en particular empleado en los vectores descritos en la presente descripción como marcador seleccionable se elige principalmente por conveniencia. Los genes marcadores seleccionables conocidos incluyen, pero no se limitan a: gen de timidina quinasa, gen de dihidrofolato reductasa, gen de xantina-guanina fosforibosil transferasa, CAD, gen de adenosina desaminasa, gen de asparagina sintetasa, numerosos genes de resistencia a antibióticos
- 15 (tetraciclina, ampicilina, kanamicina, neomicina, y similares), genes de aminoglucósido fosfotransferasa, gen de higromicina B fosfotransferasa y genes cuya expresión proporciona la presencia de un producto detectable, ya sea directa o indirectamente, como, por ejemplo, beta-galactosidasa, GFP y similares.
- 20 En algunas modalidades, los ácidos nucleicos descritos en la presente descripción comprenden al menos un gen transcripcionalmente activo, la porción del ácido nucleico comprende, además, al menos un sitio reconocido de endonucleasa de restricción, por ejemplo, un sitio de restricción que sirve como un sitio para la inserción de un ácido nucleico exógeno. Se conocen una variedad de sitios de restricción en la técnica e incluyen, pero no se limitan a: HindIII, PstI, Sall, AccI, HincII, XbaI, BamHI, SmaI, XmaI, KpnI, SacI, EcoRI y similares. En algunas modalidades, el vector incluye
- 25 un polienlazador, es decir, una serie o conjunto de sitios dispuestos cercanamente reconocidos por una pluralidad de enzimas de restricción diferentes, tales como las descritas en la presente descripción. En otras modalidades, el ácido nucleico exógeno insertado podría comprender sitios de reconocimiento de recombinasas, tales como sitios LoxP, FRT o AttB/AttP, que se reconocen por las recombinasas Cre, Flp y PhiC31, respectivamente.
- 30 En otro aspecto, la descripción se refiere a un método para generar un ácido nucleico que codifica un polipéptido específico para la unión a una secuencia de nucleótidos seleccionada, que comprende: (1) linealizar un plásmido iniciador con PspXI o nucleasa, el plásmido iniciador comprende una secuencia de nucleótidos que codifica un primer monómero que comprende un RVD específico para el primer nucleótido de la secuencia de nucleótidos seleccionada, en donde el primer monómero tiene un sitio único de PspXI o nucleasa en su extremo 3'; (2) ligar al sitio PspXI del plásmido iniciador un módulo de ADN que codifica uno o más monómeros que comprenden RVD específicos para el(los) siguiente(s)
- 35 nucleótido(s) de la secuencia de nucleótidos seleccionada, en donde el módulo de ADN tiene extremos adhesivos XhoI; y (3) repetir las etapas (1) y (2) hasta que el ácido nucleico codifique un polipéptido capaz de unirse a la secuencia de nucleótidos seleccionada. El método puede comprender, además, después de la unión, determinar la orientación del módulo de ADN en el sitio PspXI o sitio nucleasa. El método puede comprender repetir las etapas (1) y (2) de una a 30 veces.
- 40 Cuando la fuente del dominio de unión al ADN es un ácido nucleico que codifica los polipéptidos de la descripción, el ácido nucleico que codifica el polipéptido o la proteína forma parte generalmente de un módulo de expresión, como se describió anteriormente, donde los elementos adicionales proporcionan la expresión de la transposasa según sea necesario.
- 45 En algunas modalidades, los organismos multicelulares pueden prepararse mediante el uso de células mutagenizadas mediante las composiciones descritas en la presente descripción. En algunas modalidades, el organismo multicelular o unicelular es una planta o un animal no humano. En algunas modalidades, el organismo multicelular o unicelular es un vertebrado no humano. En algunas modalidades, el animal vertebrado es un mamífero no humano, tal como, por ejemplo,
- 50 un roedor (ratón o rata), ganado (cerdo, caballo, vaca, etcétera), mascotas (perro o gato) y primates no humanos.
- Los métodos descritos en la presente descripción pueden usarse en una variedad de aplicaciones en las que se desea introducir e integrar de forma estable un ácido nucleico exógeno en el genoma de una célula objetivo. *Se conocen métodos in vivo* para integrar ácido nucleico exógeno en una célula objetivo. La ruta de administración del sistema de unión a ácido nucleico al organismo multicelular o unicelular depende de varios parámetros, lo que incluye: la naturaleza de los vectores que portan los componentes del sistema, la naturaleza del vehículo de suministro, la naturaleza del organismo multicelular o unicelular y similares, donde una característica común del modo de administración es que proporciona el suministro *in vivo* de los componentes del sistema de unión al ácido nucleico a las células objetivo. En ciertas modalidades, el ADN lineal o circularizado, tal como un plásmido, se emplea como el vector para el suministro del sistema de unión al ácido nucleico a la célula objetivo. En dichas modalidades, el plásmido puede administrarse en un vehículo de suministro acuoso, tal como una solución salina. Alternativamente, puede emplearse un agente que modula la distribución del vector en el organismo multicelular o unicelular. Por ejemplo, cuando los vectores que comprenden los componentes del sistema sujeto son vectores plasmídicos, basados en lípidos tales como un liposoma, pueden emplearse vehículos, donde el vehículo basado en lípidos puede dirigirse a un tipo de célula específico para el suministro específico de células o tejidos
- 60 del vector. Alternativamente, los péptidos basados en polilisina pueden emplearse como portadores, que pueden modificarse o no con restos de direccionamiento, y similares (Brooks y otros, J. Neurosci. Methods, 1998, 80, 137-47; y
- 65

Muramatsu y otros, Int. J. Mol. Med., 1998, 1, 55-62). Los componentes del sistema pueden incorporarse, además, en vectores virales, tales como vectores derivados de adenovirus, vectores derivados de virus sindbis, vectores derivados de retrovirus, vectores híbridos y similares. Los vectores y vehículos de suministro anteriores son simplemente representativos. Puede emplearse cualquier combinación de vector/vehículo de suministro, siempre que proporcione la administración *in vivo* del sistema de unión al ácido nucleico al organismo multicelular o unicelular y a la célula objetivo.

La cantidad de ácido nucleico del vector que comprende el elemento de unión al ácido nucleico, y en muchas modalidades, la cantidad de ácido nucleico del vector que codifica el polipéptido que se introduce en la célula, es suficiente para proporcionar la escisión e inserción deseada del ácido nucleico que se une al ácido nucleico en el genoma de la célula objetivo. Como tal, la cantidad de ácido nucleico del vector introducido debería proporcionar una cantidad suficiente de actividad de unión al ADN y un número de copias suficiente del ácido nucleico que se desea insertar en la célula objetivo. La cantidad de ácido nucleico del vector que se introduce en la célula objetivo varía en dependencia de la eficiencia del protocolo de introducción particular que se emplea, como el protocolo de administración *in vivo* que se emplea.

La dosis particular de cada componente del sistema que se administra al organismo multicelular o unicelular varía en dependencia de la naturaleza del ácido nucleico que se une al ácido nucleico, por ejemplo, la naturaleza del módulo de expresión y el gen, la naturaleza del vector en el que los elementos componentes están presentes, la naturaleza del vehículo de suministro y similares. Los expertos en la técnica pueden determinar fácilmente las dosis empíricamente. Por ejemplo, en ratones donde los componentes del sistema de unión al ácido nucleico están presentes en plásmidos separados que se administran por vía intravenosa a un mamífero en un vehículo de solución salina, la cantidad de plásmido de unión al ácido nucleico que se administra en muchas modalidades varía típicamente de aproximadamente 0,5 a 40 µg y típicamente es de aproximadamente 25 µg, mientras que la cantidad de plásmido que codifica el sistema de unión al ácido nucleico que se administra típicamente varía de aproximadamente 0,5 a 25 µg y generalmente es de aproximadamente 1 µg.

Los métodos en cuestión pueden usarse para unir y producir ácidos nucleicos de diversos tamaños. Generalmente, el tamaño del ADN que se inserta en un genoma de células objetivo mediante el uso de los métodos en cuestión varía de aproximadamente 0,5 kb a 100,0 kb, generalmente de aproximadamente 1,0 kb a aproximadamente 60,0 kb, o de aproximadamente 1,0 kb a aproximadamente 10,0 kb.

La descripción puede usarse, por ejemplo, en mutagénesis de la línea germinal en una rata, ratón u otro vertebrado no humano; mutagénesis somática en una rata, ratón u otro vertebrado; transgénesis en una rata, ratón u otro vertebrado; y uso en terapia génica humana. En cada uno de estos, la composición puede suministrarse como ADN, ARN o proteína.

Las células transformadas y/o los organismos transgénicos, tales como los que contienen el ADN insertado en el ADN de la célula huésped, pueden seleccionarse de las células no transformadas y/o los organismos transformados si se incluye un marcador seleccionable como parte de las secuencias de ADN introducidas. Los marcadores seleccionables incluyen, por ejemplo, genes que proporcionan resistencia a los antibióticos; genes que modifican la fisiología del huésped, tales como, por ejemplo, la proteína verde fluorescente, para producir un fenotipo visible alterado. Las células y/u organismos que contienen estos genes son capaces de sobrevivir en presencia de concentraciones de antibióticos, insecticidas o herbicidas que matan las células/organismos no transformados o producen un fenotipo visible alterado. Mediante el uso de técnicas estándar conocidas por aquellos familiarizados con el campo, técnicas tales como, por ejemplo, la transferencia de Southern y la reacción en cadena de la polimerasa, el ADN puede aislarse a partir de células y/u organismos transgénicos para confirmar que el ADN introducido se ha insertado.

Para que la descripción se entienda de manera más eficiente, se proporcionan ejemplos más abajo. Debe entenderse que estos ejemplos son para propósitos ilustrativos solamente y no deben interpretarse como limitantes de esta invención de ninguna manera. A lo largo de estos ejemplos, las reacciones de clonación molecular y otras técnicas estándar de ADN recombinante, se llevaron a cabo de acuerdo con los métodos descritos en Maniatis y otros, Molecular Cloning - A Laboratory Manual, 2a ed., Cold Spring Harbor Press (1989)), mediante el uso de reactivos disponibles comercialmente, excepto donde se indique de cualquier otra manera.

En otros aspectos de la descripción, la descripción se refiere a vectores virales que comprenden una cualquiera o más de una secuencia de ácido nucleico descrita en la presente descripción. El vector viral se selecciona, opcionalmente, del grupo que comprende un vector retroviral, un vector adenoviral, un vector viral adenoasociado, espumavirus, un vector lentiviral y un plásmido u otro vector, tal como transposones, descritos en la solicitud. El vector retroviral comprende, opcionalmente, un vector oncoretroviral. El vector retroviral comprende, opcionalmente, un vector lentiviral.

La solicitud incluye composiciones y métodos para proporcionar una molécula de ácido nucleico codificante de RTN a un sujeto de manera que la expresión de la molécula en las células proporciona la actividad biológica del polipéptido codificado por la molécula de ácido nucleico codificante a esas células. Un ácido nucleico codificante como se usa en la presente significa un ácido nucleico que comprende nucleótidos que especifican una secuencia de aminoácidos de RTN, o una porción de esta, de la correspondiente secuencia de aminoácidos de *Ralstonia*. Una secuencia codificante puede comprender un codón de inicio y/o una secuencia de terminación.

En algunas modalidades, las composiciones de la descripción son composiciones farmacéuticas. Las composiciones farmacéuticas de esta aplicación usadas para tratar pacientes que tienen enfermedades, trastornos o estados físicos anormales podrían incluir un portador, auxiliar o excipiente aceptable.

- 5 Las composiciones farmacéuticas se administran opcionalmente por métodos *ex vivo* e *in vivo*, tales como electroporación, microinyección de ADN, suministro de ADN de liposomas y vectores de virus que tienen genomas de ARN o ADN, lo que incluye vectores de retrovirus, vectores de lentivirus, vectores de adenovirus y vectores de virus adenoasociados (AAV), virus del bosque Semliki. También son útiles los derivados o híbridos de estos vectores.
- 10 Las dosis a administrar dependen de las necesidades del paciente, del efecto deseado y de la ruta de administración elegida. Los casetes de expresión se introducen opcionalmente en las células o sus precursores mediante el uso de vehículos de suministro *ex vivo* o *in vivo* tales como liposomas o vectores de virus de ADN o ARN. Estos también se introducen opcionalmente en estas células mediante el uso de técnicas físicas tales como la microinyección o métodos químicos como la coprecipitación. Las composiciones farmacéuticas se preparan típicamente por métodos conocidos para la preparación de composiciones aceptables farmacéuticamente que se administran a pacientes, y de manera que una cantidad eficaz de la molécula de ácido nucleico se combina en una mezcla con un vehículo farmacéuticamente aceptable. Los vehículos adecuados se describen, por ejemplo, en Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., EE.UU.). Puede usarse cualquier gen marcador seleccionable.
- 15
- 20 Sobre esta base, las composiciones farmacéuticas podrían incluir un compuesto o sustancia activa, tal como una molécula de ácido nucleico, en asociación con uno o más vehículos o diluyentes aceptables farmacéuticamente, y contenidos en soluciones tamponadas con un pH adecuado e isoosmóticos con los fluidos fisiológicos. Los expertos en la técnica conocen bien los métodos para combinar los casetes de expresión con los vehículos o combinarlos con diluyentes. La composición podría incluir un agente de direccionamiento para el transporte del compuesto activo a sitios específicos dentro de las células. El casete de expresión puede comprender, además, un gen marcador seleccionable para la selección de células transformadas. Los genes marcadores seleccionables se usan para la selección de células o tejidos transformados. Los genes marcadores incluyen genes que codifican resistencia a antibióticos, tales como los que codifican neomicina fosfotransferasa II (NEO) e higromicina fosfotransferasa (HPT), así como también genes que confieren resistencia a compuestos herbicidas, tales como glufosinato de amonio, bromoxinilo, imidazolinonas y 2,4-diclorofenoxiacetato (2,4-D). Los marcadores seleccionables adicionales incluyen marcadores fenotípicos tales como la beta-galactosidasa y proteínas fluorescentes tales como la proteína verde fluorescente (GFP) (Su y otros (2004) Biotechnol Bioeng 85:610-9 y Fetter y otros (2004) Plant Cell 16:215-28), proteína cian fluorescente (CYP) (Bolte y otros (2004) J. Cell Science 117:943-54 y Kato y otros (2002) Plant Physiol 129:913-42) y proteína fluorescente amarilla (PhiYFP.TM. de Evrogen, ver, Bolte y otros (2004) J. Cell Science 117:943-54). Para marcadores seleccionables adicionales, ver generalmente, Yarranton (1992) Curr. Opin. Biotech. 3:506-511; Christopherson y otros (1992) Proc. Natl. Acad. Sci. EE.UU. 89:6314-6318; Yao y otros (1992) Cell 71:63-72; Reznikoff (1992) Mol. Microbiol. 6:2419-2422; Barkley y otros (1980) en The Operon, pp. 177-220; Hu y otros (1987) Cell 48:555-566; Brown y otros (1987) Cell 49:603-612; Figge y otros (1988) Cell 52:713-722; Deuschle y otros (1989) Proc. Natl. Acad. Sci. EE.UU. 86:5400-5404; Fuerst y otros (1989) Proc. Natl. Acad. Sci. EE.UU. 86:2549-2553; Deuschle y otros (1990) Science 248:480-483; Gossen (1993) Tesis de Doctorado, Universidad de Heidelberg; Reines y otros (1993) Proc. Natl. Acad. Sci. EE.UU. 90:1917-1921; Labow y otros (1990) Mol. Cell. Biol. 10:3343-3356; Zambretti y otros (1992) Proc. Natl Acad. Sci. EE.UU. 89:3952-3956; Baim y otros (1991) Proc. Natl. Acad. Sci. EE.UU. 88:5072-5076; Wyborski y otros (1991) Nucleic Acids Res. 19:4647-4653; Hillenand-Wissman (1989) Topics Mol. Struc. Biol. 10:143-162; Degenkolb y otros (1991) Antimicrob. Agents Chemother. 35:1591-1595; Kleinschmidt y otros (1988) Biochemistry 27:1094-1104; Bonin (1993) Tesis de Doctorado, Universidad de Heidelberg; Gossen y otros (1992) Proc. Natl. Acad. Sci. EE.UU. 89:5547-5551; Oliva y otros (1992) Antimicrob. Agents Chemother. 36:913-919; Hlavka y otros (1985) Handbook of Experimental Pharmacology, vol. 78 (Springer-Verlag, Berlín); Gill y otros (1988) Nature 334:721-724.
- 30
- 35
- 40
- 45
- 50 Se dispone de numerosos vectores de transformación de plantas y métodos para transformar plantas. Ver, por ejemplo, An, G. y otros (1986) Plant Pysiol., 81:301-305; Fry, J. y otros (1987) Plant Cell Rep. 6:321-325; Block, M. (1988) Theor. Appl Genet. 76:767-774; Hinchey y otros (1990) Stadler. Genet. Symp. 203212.203-212; Cousins y otros (1991) Aust. J. Plant Physiol. 18:481-494; Chee, P. P. y Slightom, J. L. (1992) Gene. 118:255-260; Christou y otros (1992) Trends. Biotechnol. 10:239-246; D'Halluin, y otros (1992) Bio/Technol. 10:309-314; Dhir y otros (1992) Plant Physiol. 99:81-88; Casas y otros (1993) Proc. Nat. Acad Sci. EE.UU. 90:11212-11216; Christou, P. (1993) In Vitro Cell. Dev. Biol.-Plant; 29P:119-124; Davies y otros (1993) Plant Cell Rep. 12:180-183; Dong, J. A. y Mchughen, A. (1993) Plant Sci. 91:139-148; Franklin, C. I. y Trieu, T. N. (1993) Plant. Physiol. 102:167; Golovkin y otros (1993) Plant Sci. 90:41-52; Guo Chin Sci. Bull. 38:2072-2078; Asano y otros (1994) Plant Cell Rep. 13; Ayeres N. M. y Park, W. D. (1994) Crit. Rev. Plant. Sci. 13:219-239; Barcelo y otros (1994) Plant. J. 5:583-592; Becker y otros (1994) Plant. J. 5:299-307; Borkowska y otros (1994) Acta. Physiol Plant. 16:225-230; Christou, P. (1994) Agro. Food. Ind. Hi Tech. 5: 17-27; Eapen y otros (1994) Plant Cell Rep. 13:582-586; Hartman, y otros (1994) Bio-Technology 12: 919923; Ritala y otros (1994) Plant. Mol. Biol. 24:317-325; y Wan, Y. C. y Lemaux, P. G. (1994) Plant Physiol. 104:3748.
- 55
- 60
- 65 Los métodos descritos en la presente descripción implican la introducción de un constructo polinucleotídico que comprende una secuencia de ADN en una célula huésped. Por "introducción" se pretende presentar a la planta el constructo polinucleotídico de manera que el constructo obtenga acceso al interior de la célula huésped. Los métodos no

dependen de un método en particular para introducir un constructo polinucleotídico en una célula huésped, solo que el constructo polinucleotídico obtiene acceso al interior de una célula del huésped. Los métodos para introducir constructos polinucleotídicos en bacterias, plantas, hongos y animales se conocen en la técnica, lo que incluye, pero no se limita a, métodos de transformación estable, métodos de transformación transitoria y métodos mediados por virus.

Por "transformación estable" se pretende que el constructo polinucleotídico introducido en una planta se integra en el genoma del huésped y es capaz de heredarse por la progenie de este. Por "transformación transitoria" se pretende que un constructo polinucleotídico introducido en el huésped no se integra en el genoma del huésped.

La solicitud incluye métodos y composiciones para proporcionar una molécula de ácido nucleico codificante a las células de un individuo de manera que la expresión de la molécula de ácido nucleico codificante en las células proporciona la actividad biológica o el fenotipo del polipéptido codificado por la molécula de ácido nucleico codificante. El método se refiere, además, a un método para proporcionar a un individuo que tiene una enfermedad, trastorno o estado físico anormal un polipéptido activo biológicamente mediante la administración de una molécula de ácido nucleico de la presente solicitud. El método puede realizarse *ex vivo* o *in vivo*. Los métodos y composiciones de terapia génica se demuestran, por ejemplo, en las Patentes de Estados Unidos Núms. 5,869,040, 5,639,642, 5,928,214, 5,911,983, 5,830,880, 5,910,488, 5,854,019, 5,672,344, 5,645,829, 5,741,486, 5,656,465, 5,547,932, 5,529,774, 5,436,146, 5,399,346 y 5,670,488, 5,240,846. La cantidad de polipéptido variará con las necesidades del sujeto. La dosis óptima del vector puede determinarse fácilmente mediante el uso de técnicas empíricas, por ejemplo, mediante el escalado de las dosis (ver la Patente de Estados Unidos Num. 5,910,488 para un ejemplo del escalado de las dosis). Los vectores que contienen las moléculas de ácido nucleico de la aplicación se administran típicamente a mamíferos, preferentemente, humanos, en terapia génica mediante el uso de las técnicas que se describen más abajo. Los polipéptidos producidos a partir de las moléculas de ácido nucleico también se administran opcionalmente a mamíferos, preferentemente, seres humanos. La solicitud se refiere a un método para el tratamiento médico de un mamífero que lo necesita, preferentemente, un ser humano, mediante la administración al mamífero de un vector de la aplicación o una célula que contiene un vector de la aplicación. A un receptor, preferentemente, un ser humano, que desarrolla un evento adverso, tal como la enfermedad de injerto contra huésped, se le administra típicamente un medicamento, tal como AZT, que es un sustrato para las moléculas *tmpr* modificadas de la aplicación. Las enfermedades, tales como las enfermedades de la sangre o las enfermedades neuronales (neurodegenerativas), que se tratan fácilmente se describen en esta solicitud y se conocen en la técnica (por ejemplo, enfermedades tales como la talasemia o la anemia de células falciformes que se tratan mediante la administración de un gen de globina como se describe en la solicitud de patente de Canadá núm. 2,246,005). Las enfermedades de la sangre que pueden tratarse mediante trasplante de células madre incluyen leucemias, síndromes mielodisplásicos, trastornos de células madre, trastornos mieloproliferativos, trastornos linfoproliferativos, trastornos de fagocitos, trastornos metabólicos hereditarios, trastornos histiocíticos, alteraciones hereditarias de eritrocitos, trastornos hereditarios del sistema inmunitario, anomalías hereditarias de plaquetas, trastornos de células plasmáticas, neoplasias malignas (ver también, Medical Professional's Guide to Unrelated Donor Stem Cell Transplants, 4a edición). Las enfermedades de los nervios de células madre a tratar mediante trasplante de células madre neurales incluyen enfermedades que provocan daño o pérdida de las células neurales, por ejemplo, parálisis, enfermedad de Parkinson, enfermedad de Alzheimer, ALS, esclerosis múltiple). El vector de la aplicación es útil como marcador de células madre y para expresar genes que provocan que las células madre se diferencien (por ejemplo, factor de crecimiento).

Pueden usarse diversos enfoques para la terapia génica. La solicitud incluye un proceso para proporcionar a un ser humano un polipéptido terapéutico, lo que incluye: introducir células humanas en un humano, dichas células humanas se han tratado *in vitro* o *ex vivo* para insertar en ellas un vector de la solicitud, las células humanas expresan *in vivo* en dicho ser humano una cantidad terapéuticamente eficaz de dicho polipéptido terapéutico.

El método se refiere, además, a un método para producir una reserva de virus recombinante mediante la producción de virus adecuados para la terapia génica que comprende ADN modificado que codifica la globina. Este método implica, preferentemente, transfectar células permisivas para la replicación del virus (el virus que contiene la globina modificada) y recolectar el virus producido.

Puede emplearse cotransfección (ADN y marcador en moléculas separadas) (ver, por ejemplo, la Patente de Estados Unidos Núm. 5,928,914 y la patente de Estados Unidos Núm. 5,817,492). Además, puede usarse un casete o marcador de detección (tal como un marcador de proteína fluorescente verde o un derivado, CD19 o CD25) dentro del propio vector (preferentemente, un vector viral).

Los métodos de la descripción pueden usarse para mutar cualquier célula madre eucariota, lo que incluye, pero no se limita a, haploide, diploide, triploide, tetraploide o aneuploide. En una modalidad, la célula es diploide. Las células madre en las que los métodos descritos en la presente descripción pueden usarse ventajosamente incluyen, pero no se limitan a, células madre tales como células madre somáticas, SSC, células ES, células iPS, embriones no humanos o cualquier célula capaz de desarrollarse en uno o más organismos.

En una modalidad, la descripción se refiere a un método para producir una célula madre con genes anulados, introducidos o de cualquier otra manera modificada genéticamente específica de sitio. La mutación específica de sitio se genera mediante el uso de un RTN que escinde el sitio deseado, seguido de NHEJ, lo que resulta en mutaciones de delección. La mutación específica de sitio puede producirse en células madre espermatozonales no humanas (SSC) que se usan para generar organismos heterocigotos u homocigotos modificados genéticamente.

- En otra modalidad, la descripción se refiere a un método para producir una célula madre con genes anulados, introducidos o de cualquier otra manera modificada genéticamente específica de sitio. La mutación específica de sitio se genera mediante el uso de un RTN que escinde el sitio deseado lo que resulta en mutaciones de delección. La mutación específica de sitio se produce en células madre embrionarias (ES), que se usan para generar organismos heterocigotos u homocigotos modificados genéticamente.
- En otra modalidad, la descripción comprende métodos para producir una célula madre con genes anulados, introducidos o de cualquier otra manera modificada genéticamente específica de sitio. La mutación específica de sitio se genera mediante el uso de un RTN que escinde el sitio deseado lo que resulta en mutaciones de delección. La mutación específica de sitio se produce en células madre pluripotentes inducidas (iPS), que se usan para generar organismos heterocigotos u homocigotos modificados genéticamente.
- En otra modalidad, la descripción comprende métodos para producir una célula madre con genes anulados, introducidos o de cualquier otra manera modificada genéticamente específica de sitio. La mutación específica de sitio se genera mediante el uso de un RTN que escinde el sitio deseado lo que resulta en mutaciones de delección. La mutación específica de sitio se produce en embriones no humanos que se usan para generar organismos heterocigotos u homocigotos modificados genéticamente.
- En ciertas modalidades, las células pueden mutar dentro del organismo o dentro del entorno nativo como en los explantes de tejido (por ejemplo, *in vivo* o *in situ*). Alternativamente, los tejidos o las células madre aisladas del organismo mediante el uso de métodos conocidos en la técnica y los genes pueden mutarse de acuerdo con los métodos presentes. Los tejidos o las células madre se mantienen ya sea en cultivo (por ejemplo, *in vitro*), o reimplantados en un tejido u organismo (por ejemplo, *ex vivo*).
- Los RTN comprenden la función de la proteína efectora, tal como una nucleasa.
- En algunas modalidades, la descripción se refiere a composiciones que comprenden uno cualquiera de los ácidos nucleicos o polipéptidos o fragmentos de estos descritos en el Ejemplo 2.
- En algunas modalidades, la descripción se refiere a composiciones que comprenden uno cualquiera de los ácidos nucleicos o polipéptidos descritos en la presente descripción.
- En algunas modalidades de la descripción, la modificación genética de las SSC mediante el uso de RTN se relaciona con la generación de mutaciones múltiples en SSC o líneas de SSC separadas seguida de la agrupación o combinación de las SSC o líneas de SSC separadas y su inyección en un solo receptor masculino, lo que se relaciona con la generación de múltiples organismos modificados genéticamente que contienen una o más mutaciones en menos etapas experimentales y en un período de tiempo más corto de lo que es posible con otros sistemas. Las células madre o líneas de células madre separadas pueden ser quince o más. En alguna modalidad, la modificación genética de las células madre mediante el uso de RTN se relaciona con la generación de mutaciones múltiples en células madre o líneas de células madre separadas seguida de la agrupación o combinación de células madre o líneas de células madre separadas y su inyección en un solo receptor masculino, lo que se relaciona con la generación de múltiples organismos modificados genéticamente que contienen una o más mutaciones en menos etapas experimentales y en un período de tiempo más corto de lo que es posible con otros sistemas. Las células madre o líneas de células madre separadas pueden ser quince o más.
- En algunas modalidades, aumentar el número de agrupaciones o líneas distintas o separadas de SSC modificadas genéticamente o células madre modificadas o células embrionarias no humanas, que pueden usarse para generar un organismo modificado genéticamente, no aumenta la cantidad de esfuerzo, tiempo, y los recursos usados, así como tampoco disminuye la eficiencia de la producción de organismos modificados genéticamente. Las múltiples SSC modificadas genéticamente separadas y distintas pueden trasplantarse en un solo receptor estéril. La población mixta de células modificadas genéticamente distintas (SSC o células madre), que se derivan de agrupaciones de células separadas de dos o más agrupaciones a quince o más agrupaciones maduras dentro del receptor estéril. El receptor estéril se reproduce a continuación con múltiples hembras de tipo salvaje que pueden ser dos o más, tres o más, cuatro o más, cinco o más, seis o más, siete o más, ocho o más, nueve o más, diez o más, once o más, doce o más, trece o más, catorce o más, quince o más, dieciséis o más, diecisiete o más, dieciocho o más, diecinueve o más, veinte o más. Estas múltiples hembras producen crías que han incorporado la mutación deseada en su línea germinal.
- En algunas modalidades, aumentar el número de agrupaciones o líneas distintas o separadas de células modificadas genéticamente, que pueden usarse para generar un organismo modificado genéticamente, no aumenta la cantidad de esfuerzo, tiempo y recursos usados, así como tampoco disminuye la eficiencia de la producción de organismos modificados genéticamente. La rata receptora estéril puede ser receptora de múltiples rondas de células modificadas genéticamente separadas o distintas. La rata estéril puede ser receptora de quince o más células modificadas genéticamente diferentes y reproducirse con veinte o más hembras de tipo salvaje para producir quince o más organismos modificados genéticamente por separado. Después de la primera ronda de reproducción, el macho estéril puede tratarse para eliminar la primera ronda de células modificadas genéticamente y convertirse en un receptor de otra ronda de quince

o más células modificadas genéticamente por separado o distintas, reproducirse con veinte o más hembras de tipo salvaje para producir quince o más organismos modificados genéticamente separados. El macho estéril puede ser un receptor de poblaciones mixtas de quince o más células modificadas genéticamente y reproducirse con veinte o más hembras de tipo salvaje dos veces o más, tres veces o más, cuatro veces o más, o cinco veces o más.

5 En algunas modalidades, aumentar el número de agrupaciones o líneas distintas o separadas de células modificadas genéticamente, que pueden usarse para generar un organismo modificado genéticamente, no aumenta la cantidad de esfuerzo, tiempo y recursos usados, así como tampoco disminuye la eficiencia de la producción de organismos modificados genéticamente. Aumentar el número de células modificadas genéticamente no requiere el esfuerzo y los
10 recursos de otros sistemas celulares, tales como las células madre embrionarias (ES) o los embriones no humanos. El aumento de la cantidad de células ES modificadas genéticamente para la producción de organismos modificados genéticamente requiere un aumento en el número de etapas técnicas, tales como las inyecciones de blastocitos, así como también el número de cirugías de transferencia de oviductos. En algunas modalidades, el método no comprende inyección de blastocitos, transferencia de oviducto, microinyección de ADN, reimplantación de cigotos inyectados o reproducción de
15 progenie quimérica. El sistema celular puede producir quince o más poblaciones separadas de células madre modificadas genéticamente para la producción de organismos modificados genéticamente en una sola etapa, mientras que para producir quince o más células ES modificadas genéticamente por separado, deben realizarse quince o más etapas separadas en todos los niveles del procedimiento, que incluye, pero no se limita a, inyección de blastocitos, transferencia de oviducto, producción de cigoto, preparación de ADN, microinyección de ADN, reimplantación de cigotos inyectados o
20 reproducción de progenie quimérica.

En algunas modalidades, la modificación genética de las SSC mediante el uso de RTN se relaciona con la generación de organismos modificados genéticamente sin requerir las etapas necesarias para producir organismos modificados genéticamente a partir de células madre alternativas, lo que incluye, pero no se limita a, células madre embrionarias,
25 embriones no humanos, células madre pluripotentes inducidas (iPS), células madre somáticas. La modificación genética en células madre alternativas incluye, pero no se limita a, la producción de cigotos, la preparación de ADN, la microinyección de ADN, el reimplante de cigotos inyectados o la reproducción de la progenie quimérica.

En algunas modalidades, las células madre descritas en la presente descripción comprenden uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón de variantes de estas. En algunas modalidades, las células madre comprenden uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón de variantes derivadas de las secuencias de la Tabla 2.

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos un 70 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones en tándem invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos un 75 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones en tándem invertidas (ITR) de un transposón en donde las repeticiones terminales invertidas de secuencia variante son al menos un 80 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos un 85 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos 90 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos 95 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos un 96 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos 97 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos 98 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

- 5 En algunas modalidades, la descripción comprende uno o más transposones, una o más repeticiones terminales invertidas (ITR) de un transposón en donde las repeticiones en tándem invertidas de secuencia variante son al menos 99 % homólogas a las ITR conocidas y elementos de un transposón conocidos (mostrados en la Tabla 2).

TABLA 2

10	ITR de un transposón
	Sleeping Beauty
	Repetición en tándem invertida 5':
15	CAGTTGAAGTCGGAAGTTTACATACACTTAAGTTGGAGTCATTAAAACTCGTTTT
	TCAACTACTCCACAAATTTCTTGTTAACAAACAATAGTTTTGGCAAGTCAGTTAG
	GACATCTACTTTGTGCATGACACAAGTCATTTTTCCAACAATTGTTTACAGACAG
20	ATTATTTCACTTATAATTCACGTGTATCACAATCCAGTGGGTCAGAAGTTTACATA
	CACTAAGT
25	Repetición en tándem invertida 3':
	ATTGAGTGTATGTAACTTCTGACCCACTGGGAATGTGATGAAAGAAATAAAAG
	CTGAAATGAATCATTCTCTCTACTATTATTCTGATATTTACATTCTTAAATAAA
30	GTGGTGATCCTAACTGACCTAAGACAGGGAATTTTTACTAGGATTAAATGTCAGG
	AATTGTGAAAAAGTGAGTTTAAATGTATTTGGCTAAGGTGTATGTAACTTCCGA
35	CTTCAACTG
	piggyBac
	Repetición en tándem invertida 5':
40	CCCTAGAAAGATAGTCTGCGTAAATTTGACGCATGCATTCTTGAAATATTGCTCT
	CTCTTTCTAAATAGCGCGAATCCGTCGCTGTGCATTTAGGACATCTCAGTCGCCG
	CTTGAGCTCCCGTGAGGCGTGCTTGTCATGCGGTAAGTGTCAGTATTTGAA
45	CTATAACGACCGCGTGAGTCAAAATGACGCATGATTATCTTTTACGTGACTTTTA
	AGATTTAACTCATACGATAATTATATTGTTATTTTCATGTTCTACTTACGTGATAAC
	TTATTATATATATATTTTCTTGTTATAGATATC (secuencia mínima subrayada y en
50	negritas, es decir, primer 35 bp)
	Repetición en tándem invertida 3':
55	TAAAAGTTTTGTTACTTTATAGAAGAAATTTTGAGTTTTTGTTTTTTTTTAATAAA
	TAAATAAACATAAATAAATTGTTTGTTGAATTTATTATTAGTATGTAAGTGTAAG
	TATAATAAACTTAATATCTATTCAAATTAATAAATAAACCTCGATATACAGACC
60	GATAAAACACATGCGTCAATTTTACGCATGATTATCTTTAACGTACGTCACAATA
	TGATTATCTTTCTAGGG (secuencia mínima subrayada y en negritas, es decir, primer 35 bp)

La descripción comprende, además, ADN que codifica para uno cualquiera de los polipéptidos descritos en la presente descripción con cualquier número de unidades de repetición descritas en la presente descripción.

5 Cualesquiera polipéptidos de la descripción o ácidos nucleicos que codifican estos polipéptidos puede usarse en los métodos descritos en la Solicitud PCT Núm. PCT/IB2010/000154.

En un aspecto de la descripción, se proporciona una composición que comprende una o más células modificadas mediante uno o más polipéptidos descritos en la presente descripción. En algunas modalidades, la composición comprende una o más células madre. En algunas modalidades, la composición comprende una o más células madre de mamífero
10 modificadas mediante uno o más polipéptidos descritos en la presente descripción. En algunas modalidades, la composición comprende una o más células iPSC modificadas mediante uno o más polipéptidos descritos en la presente descripción. En algunas modalidades, la composición comprende una o más células madre humanas modificadas mediante uno o más polipéptidos descritos en la presente descripción. En algunas modalidades, la composición comprende una o más células madre espermatozonales no humanas modificadas mediante uno o más polipéptidos
15 descritos en la presente descripción. En algunas modalidades, la célula se deriva de un mamífero. En algunas modalidades, la célula es de una rata o un minicerdo. En algunas modalidades, el mamífero es una rata macho estéril o un minicerdo macho estéril. En algunas modalidades, la rata o el minicerdo es DAZL deficiente o DAZL^{-/-}. En algunas modalidades, la descripción se refiere a una colonia de organismos modificados genéticamente que comprende:

20 al menos un organismo que comprende una o más células madre, la una o más células madre comprenden una o más de las siguientes mutaciones: (i) una mutación de delección; (ii) una mutación de anulación; y/o (iii) una adición de una secuencia de ácido nucleico heterólogo; la una o más mutaciones de (i), (ii) y/o (iii) son mutaciones específicas de sitio provocadas mediante uno o más polipéptidos descritos en la presente descripción (uno o más RTN); y (b) la progenie del organismo de la subparte (a).

25 En algunas modalidades, la célula o animal transgénico no humano, colonia o progenie de esta comprende una secuencia de ácido nucleico heterólogo que comprende un marcador seleccionable o un gen ortólogo. En algunas modalidades, el al menos un organismo y la progenie comprenden además al menos una repetición terminal invertida de un transposón o variante de este.

30 En algunas modalidades, el al menos un organismo y la progenie comprenden además un ácido nucleico que comprende una secuencia de ácido nucleico que es al menos 70 % homólogo a cualquiera o una combinación de: SEQ ID NO:1 a 19, o cualquier secuencia de la Tabla 1, o cualesquiera variantes o fragmentos funcionales de esta. En algunas modalidades, la descripción se refiere a un método para generar uno o más organismos modificados genéticamente que comprende: (a) poner en contacto al menos una célula madre derivada del linaje de la línea germinal de un animal no humano o planta por la célula madre con: (i) al menos un RTN que muta un gen de interés; o (ii) al menos un vector de expresión que codifica un RTN que muta un gen de interés, lo que crea de esta manera al menos una célula madre que comprende al menos una mutación en un gen de interés; (b) expandir un cultivo *in vitro* de al menos una célula madre que comprende al menos una mutación en un gen de interés; (c) implantar una o más células madre del cultivo de la etapa
40 (b) en un organismo.

En algunas modalidades, la descripción se refiere a un método para generar uno o más organismos modificados genéticamente que comprende: (a) poner en contacto al menos un primer y un segundo conjunto de células madre derivadas del linaje de la línea germinal de un animal no humano o planta con: (i) al menos un RTN que muta un gen de interés; o (ii) al menos un vector de expresión que codifica un RTN que muta un gen de interés, lo que crea de esta manera al menos un primer y un segundo conjunto de células madre que comprenden al menos una mutación en un gen de interés; (b) expandir un cultivo *in vitro* de al menos una célula madre que comprende al menos una mutación en un gen de interés;

50 (c) implantar uno o más conjuntos de células madre a partir del cultivo de la etapa (b) en un organismo. En algunas modalidades, el método comprende además un tercer, cuarto, quinto, sexto, séptimo, octavo, noveno o diez o más conjuntos de células madre que se han mutado de una manera específica de sitio por un RTN y, en cuyo caso, después de expandir cada uno del tercer, cuarto, quinto, sexto, séptimo, octavo, noveno o diez o más conjuntos de células madre mutadas, cada conjunto se trasplanta en un solo organismo. En algunas modalidades, el organismo individual que comprende un conjunto de células madre mutadas es un hombre estéril.

En algunas modalidades de la descripción, el organismo es capaz de transmitir al menos una mutación en un gen de interés a la progenie mediante transmisión por la línea germinal. En algunas modalidades, el organismo modificado genéticamente es un mamífero no humano. En algunas modalidades, el organismo modificado genéticamente es una rata o un minicerdo. En algunas modalidades, el organismo modificado genéticamente es una rata macho estéril o un minicerdo macho estéril.

60 Descrito allí, el método comprende además: reproducir el organismo implantado con la una o más células madre con otro animal no humano para generar una o más progenie que comprenda el gen mutado de interés. Descrito allí, el método comprende, además: reproducir el organismo implantado con uno o más conjuntos de células madre con otro animal no

humano para generar una o más progenies que comprendan uno o más genes mutados de interés que correspondan a cada una de las líneas de células madre mutantes.

En algunas modalidades, la progenie son mamíferos.

5

En algunas modalidades de la descripción, un método de reproducción de una colonia de organismos modificados genéticamente comprende:

10

(a) poner en contacto al menos una célula madre derivada del linaje de la línea germinal de un animal no humano o planta por la célula madre con: (i) al menos un RTN que muta un gen de interés; o (ii) al menos un vector de expresión que codifica un RTN que muta un gen de interés, lo que crea de esta manera una célula madre que comprende al menos una mutación en un gen de interés;

15

(b) expandir un cultivo *in vitro* de la célula madre que comprende al menos una mutación en un gen de interés;

(c) implantar la al menos una célula madre que comprende al menos una mutación en un gen de interés del cultivo de la etapa (b) en un primer organismo.

20

(d) reproducir el primer organismo con un segundo organismo de la misma especie;

(e) seleccionar la progenie del primer y segundo organismo que comprende al menos una mutación en un gen de interés; y

25

(f) reproducir la progenie para crear una colonia de organismos que comprenda la al menos una mutación en un gen de interés.

En algunas modalidades, el primer y el segundo organismo son mamíferos.

En algunas modalidades, el primer y el segundo organismo son ratas o minicerdos.

30

En algunas modalidades, la descripción se refiere a un método de fabricación de una primera generación filial de organismos modificados genéticamente que comprende dos o más subconjuntos distintos de organismos, el método comprende:

35

(a) poner en contacto una primera célula madre con: (i) un RTN que muta un primer gen de interés; o (ii) un vector de expresión que codifica un RTN que muta un primer gen de interés; lo que crea de esta manera una primera célula madre que comprende una primera mutación;

40

(b) poner en contacto una segunda célula madre con un agente modificador, lo que crea de esta manera una segunda célula madre que comprende una segunda mutación;

(c) expandir un cultivo *in vitro* de cada una de las células madre primera y segunda;

45

(d) implantar una población mixta de células madre que comprende la primera y la segunda células madre en un organismo;

(e) reproducir el organismo con otro organismo de la misma especie.

50

En algunas modalidades, la primera generación filial de organismos modificados genéticamente comprende dos o más conjuntos de organismos, cada conjunto comprende una mutación de interés distinta derivada a partir de un haplotipo de células madre distintas trasplantadas en un progenitor del organismo.

En algunas modalidades, al menos una célula madre de la población mixta es una célula madre espermatogonial no humana de un mamífero.

55

En algunas modalidades, el organismo es un mamífero.

En algunas modalidades, el kit comprende:

60

(a) un RTN o una secuencia de ácido nucleico que codifica un RTN que escinde una secuencia de ácido nucleico en un gen de interés; y

(b) un manual de instrucciones que comprende instrucciones; y, opcionalmente

65

En algunas modalidades, el kit comprende:

(a) En algunas modalidades; y, opcionalmente

(b) medios de cultivo para la una o más células madre o uno o más embriones no humanos. En algunas modalidades, el kit comprende:

(a) un RTN o una secuencia de ácido nucleico que codifica un RTN que escinde una secuencia de ácido nucleico en un gen de interés; y opcionalmente

(b) medios de cultivo para la una o más células madre o uno o más embriones no humanos.

En algunas modalidades, el kit comprende:

(a) un RTN o una secuencia de ácido nucleico que codifica un RTN que escinde una secuencia de ácido nucleico en un gen de interés; y

(b) una o más líneas de células madre derivadas de un linaje de línea germinal no humana de animal o planta; y, opcionalmente

(c) medios de cultivo para la una o más células madre o uno o más embriones no humanos; y, opcionalmente

(d) un manual de instrucciones que comprende instrucciones sobre cómo mutar la una o más células madre con el RTN o una secuencia de ácido nucleico que codifica el RTN que escinde una secuencia de ácido nucleico en un gen de interés.

Tabla 1

Isteqvvaia snkggkqale avkahlldl gapyv
Isteqvvaia snkggkqale aveaqllrlr aapye
Isteqvvaia snkggkqale avgaqllalr avpya
Isteqvvaia snkggkqale avgaqllalr avpye
Isteqvvaia snkggkqale avkahlldl gapyv
Isteqvvaia snkggkqale avkahlldl gapyv
Isteqvvaia snkggkqale avkahlldl gapyv
Isteqvvaia snkggkqale avkaqllelr gapya
Isteqvvaia snkggkqale aveaqllrlr aapye
Isteqvvaia snkggkqale avkahlldl gapyv
Isteqvvaia snkggkqale avkahlldl gapyv
Isteqvvaia snnggkqale avkaqlldlr gapya
Isteqvvaia snnggkqale avkaqlpvlr rapyg
Ispeqvvaia snnggkpale avkaqllelr aapye
Ispeqvvaia snnggkpale avkalllrlr aapye
Isteqvvaia snnggkpale avkalllelr aapye
Isteqvvaia snnggkqale avktqllalr tapye
Isteqvvaia snnggkqale avkaqlpalr aapye
Isteqvvaia snnggkqale avkaqlvlr aapyg
Isteqvvaia annggkqale avrallpvlr vapye
Ispeqvvaia snnggkqale avkaqllelr aapye
Isteqvvaia snnggkqale avkaqllelr aapye
Ispeqvvaia snnggkqale avkaqllelr aapye
Isteqvvaia snnggkqale avkaqllelr aapye
Ispeqvvaia snnggkqale avkaqllelr aapye
Isteqvvaia snnggkqale avkaqllelr aapye

5	Isteqvvaia snnggkqale avkalllelr aapye
	Isteqvvaia snnggkqale avktqlalr tapye
	Isteqvvaia snnggkqale avkaqlpalr aapye
	Ispeqvvaia snnggkqale avrallpvlr vapye
10	Istaqvvaia sngggkqale gigeqlklr tapyg
	Isteqvvaia sngggkqale gigkqlqelr aaphg
	Istgqvvaia sngggrqale avreqlalr avpye
	Istgqvvaia sngggrqale avreqlalr avpye
15	Istaqvvaia sngggkqale gigeqlklr tapyg
	Istaqvvaia sngggkqale gigeqlklr tapyg
	Istaqvvaia sngggkqale gigeqlklr tapyg
20	Istaqvvaia sngggkqale gigeqlklr tapyg
	Isteqvvaia sngggkqale gigkqlqelr aaphg
	Intaqvvaia shdggkpale avwaklpvlr gvpye
25	Intaqvvaia shdggkpale avwaklpvlr gvpye
	Intaqvvaia shdggkpale avwaklpvlr gvpya
	Isteqvvaia shdggkpale avgaqlvalr aapya
	Istaqvvaia shdggkpale avgtqlvalr aapya
30	Isteqvvaia shdggkpale avgaqlvalr aapya
	Intaqivaia shdggkpale avwaklpvlr gapya
	Istaqvvaia shdggkpale avrkqlpvlr gvphq
35	Istaqvvaia shdggkpale avwaklpvlr gapya
	Intaqvvaia shdggkpale avwaklpvlr gvpye
	Intaqvvaia shdggkpale avwaklpvlr gvpya
40	Istaqvvaia shdggkpale avgaqlvelr aapya
	Isteqvvaia shdggkpale avgaqlvalr aapya
	Intaqvvaia shdggkpale avraklpvlr gvpya
45	Itpqqvvaia shdggkpale avwaklpvlr gvpya
	Itpqqvvaia shdggkpale avwaklpvlr gvpya
	Istaqvvaia ssiggrqale avkvqlpvlr aapyg
50	Isteqvvaia nsigggkqale avkvqlpvlr aapye
	Isteqvvaia nsigggkqale avkvqlpvlr aapye
	Istaqvvaia ssiggrqale avkvqlpvlr aapyg

55 EJEMPLOS

Ejemplo 1: Generar vectores de ácido nucleico con TAL de *Ralstonia* (RTAL) con análisis funcional

60 El análisis de conglomerado y revisión de homologías de secuencia de genoma de *Ralstonia* reveló la secuencia de SEQ ID NO:1 que es homóloga a las secuencias TAL conocidas.

65 Las secuencias de ácido nucleico que codifican los polipéptidos descritos en la presente descripción se crearán a través de técnicas de biología molecular conocidas por los expertos en la técnica. Las secuencias de ADN, por ejemplo, se sintetizarán con los sitios de restricción XbaI y/o SalI que flanquean cada secuencia de ácido nucleico que codifica las poliproteínas de la descripción. Se realizará la reacción en cadena de la polimerasa para amplificar el ADN con ciertos sitios de endonucleasa de restricción. Las secuencias se purificarán en gel, se aislarán y se reconstituirán en agua o

tampón adecuado para las reacciones de unión. Un plásmido que codifica una proteína con función efectora (tal como la función nucleasa) que comprende los elementos reguladores necesarios se unirá a una o más de las secuencias de ácido nucleico que codifican las siguientes secuencias en los sitios de clonación múltiples del plásmido:

- 5 a. LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV
- b. LSTEQVVAIAS NN GGKQALEAVKAQLLELRAAPYE
- 10 c. LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
- d. LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPIYA
- e. LSTEQVVTIAS SI GGKQALEAVKVQLPVLRAAPYE

15 Las secuencias de plásmidos se transformarán en bacterias adecuadas para la producción de elevados números de copias de plásmidos. Los plásmidos que contienen al menos un polipéptido anterior pueden seleccionarse mediante el uso de selección con antibióticos, aislarse y purificarse a partir de células bacterianas mediante el uso de técnicas conocidas por los expertos en la técnica.

20 Además, se construirán plásmidos y se validarán las pruebas in vitro de polipéptidos de unión al ADN expresados mediante el uso de los métodos descritos en Nature Biotechnology 2012 Mayo; 30(5):460-5. "FLASH assembly of TALENs for high-throughput genome editing". Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK.

Construcción de un archivo de plásmidos que codifican repeticiones de TALE preensambladas

25 Intentamos construir matrices de repeticiones de TALE mediante el uso de la misma arquitectura descrita por primera vez por Miller, J.C. y otros "A TALE nuclease architecture for efficient genome editing." Nat Biotechnol. 2011; 29:143-148 en la que las cadenas principales de repeticiones de TALE que difieren ligeramente en sus secuencias de aminoácidos y ADN se producen en un patrón repetido. En algunas modalidades, designamos la primera repetición de TALE amino-

30 terminal en un conjunto como la unidad α . Esto se sigue por las unidades β , γ y δ y después una unidad ϵ que es esencialmente idéntica a la unidad α , excepto por el posicionamiento diferente de un sitio de restricción Tipo IIS en el extremo 5' (requerido para permitir la creación de un saliente único en la unidad α necesario para la clonación). La unidad ϵ se sigue nuevamente por repeticiones de unidades β , γ , δ y ϵ . Debido a las restricciones relacionadas con la creación de un extremo 3' requerido para la clonación, se requirieron secuencias de ADN ligeramente modificadas para las matrices de repeticiones de TALE que terminan con una unidad γ o ϵ carboxiterminal.

35 Preparación de fragmentos de ADN que codifican repeticiones de TALE para ensamblaje mediante FLASH

Para preparar fragmentos de ADN que codifican unidades α para su uso en el ensamblaje mediante FLASH, realizaremos 20 rondas de PCR con cada plásmido de unidad α como plantilla mediante el uso de los cebadores oJS2581(5'-Biotina-TCTAGAGAAGACAAGAACCTGACC-3') y oJS2582(5'-GGATCCGGTCTCTTAAGGCCGTGG-3'). Los productos de PCR

40 resultantes se biotinilarán en el extremo 5'. A continuación, cada producto de PCR α se digerirá con 40 unidades de enzima de restricción BsaI-HF para generar salientes de 4 pb, purificados mediante el uso del kit de purificación de PCR QIAquick (QIAGEN) de acuerdo con las instrucciones del fabricante, excepto que el producto final se eluirá en 50 μ l de EB 0,1 X.

45 Para preparar fragmentos de ADN que codifican repeticiones de polipéptidos, digeriremos 10 μ g de cada uno de estos plásmidos con 50 unidades de enzima de restricción BbsI en NEBuffer 2 durante 2 horas a 37 °C seguido de digestiones de restricción en serie realizados en NEBuffer 4 a 37 °C mediante el uso de 100 unidades de cada una de las enzimas XbaI, BamHI-HF y Sall-HF que se añadirán a intervalos de 5 minutos. El último conjunto de digestiones de restricción se

50 diseñará para escindir la cadena principal del plásmido para garantizar que este fragmento de ADN más grande no interfiera con las ligaduras posteriores realizadas durante el proceso de ensamblaje mediante FLASH. Estas reacciones de digestión de restricción se purificarán mediante el uso del kit de purificación por PCR QIAquick (QIAGEN) de acuerdo con las instrucciones del fabricante, excepto que el producto final se eluirá en 180 μ l de EB 0,1 X.

55 Ensamblaje mediante FLASH automatizado

Todas las etapas del ensamblaje por FLASH se realizarán mediante el uso de una estación de trabajo para manejo de líquidos Sciclone G3 (Caliper) o un dispositivo similar vendido por otra compañía en placas de 96 pocillos y mediante el uso de un imán SPRIplate de 96 anillos (Beckman Coulter Genomics) y un imán lateral DynaMag-96 (Life Technologies).

60 En la primera etapa de FLASH, un fragmento de unidad α biotinilada se unirá al primer fragmento $\beta\gamma\delta\epsilon$ y después los fragmentos $\alpha\beta\gamma\delta\epsilon$ resultantes se unirán a perlas magnéticas recubiertas de estreptavidina Dynabeads MyOne C1 (Life Technologies) en tampón B&W 2 X. A continuación, las perlas se atraerán hacia el lado del pocillo al colocar la placa en el imán y después se lavarán con 100 μ l de tampón B&W con Tween 20 al 0,005 % (Sigma) y nuevamente con 100 μ l de albúmina de suero bovino 0,1 mg/ml (BSA) (New England Biolabs). Se unirán fragmentos $\beta\gamma\delta\epsilon$ adicionales al retirar la

65 placa del imán, resuspender las perlas en solución en cada pocillo, digerir el fragmento unido a la perla con la enzima de restricción BsaI-HF, colocar la placa en el imán, lavar con 100 μ l de B&W/Tween20 seguido de 100 μ l de BSA 0,1 mg/ml,

y después unir el siguiente fragmento. Este proceso se repetirá varias veces con unidades $\beta\gamma\delta\epsilon$ adicionales para extender el fragmento unido a la perla. El último fragmento a unir siempre es una unidad β , $\beta\gamma^*$, $\beta\gamma\delta$ o $\delta\epsilon^*$ para permitir la clonación del fragmento de longitud completa en vectores de expresión (tenga en cuenta que los fragmentos que terminan con una unidad $\delta\epsilon^*$ siempre irán precedidos de la unión de una unidad $\beta\gamma$).

El fragmento final unido a la perla de longitud completa se digerirá con 40 unidades de enzima de restricción BsaI-HF seguido de 25 unidades de enzima de restricción BbsI (New England Biolabs). La digestión con BbsI liberará el fragmento de las perlas y generará un saliente único 5' para la clonación del fragmento. La digestión con BsaI-HF resulta en la creación de un saliente único 3' para la clonación.

Subclonación de fragmentos de ADN que codifican el conjunto de repeticiones TALE en vectores de expresión TALEN

Subclonaremos fragmentos de ADN que codifican nuestras matrices de repeticiones TALE ensambladas mediante FLASH en vectores de expresión TALE. En algunas modalidades, habrá 4 o más plásmidos separados. En algunas modalidades, los vectores incluirán un promotor de CMV, un codón de inicio de traducción optimizado para la expresión de células de mamífero, una etiqueta de epítipo FLAG triple, una señal de localización nuclear, aminoácidos 153 a 288 de la proteína TALE 13 (según se numeraron por Miller y otros 6), dos sitios de restricción de tipo IIS BsmBI únicos y posicionados cercanamente, un dominio de repetición TALE 0,5 que codifica los RVD, aminoácidos 715 a 777 de la proteína TALE 13, y el dominio de escisión FokI de tipo salvaje.

Todos los fragmentos de ADN ensamblados mediante FLASH poseerán salientes que permitirán la clonación direccional en cualquiera de los vectores de expresión que se digerirán con BsmBI. Los vectores de expresión TALEN estándar (cada uno con una repetición TALE 0,5 diferente) están disponibles de proveedores tales como Addgene y las secuencias completas de estos plásmidos están disponibles gratuitamente en una página web dedicada a estos constructos: <http://www.addgene.org/talengineering/ expressionvectors/> para la construcción sintética.

Para preparar un vector de expresión TALEN para la subclonación, digeriremos 5 μ g de ADN plasmídico con 50 unidades de enzima de restricción BsmBI (New England Biolabs) en NEBuffer 3 durante 8 horas a 55 grados C. El ADN digerido se purificará mediante el uso de 90 μ l de perlas Ampure XP (Agencourt) de acuerdo con las instrucciones del fabricante y se diluirán a una concentración final de 5 ng/ μ l en TrisHCl 1 mM. Los conjuntos de repeticiones TALE ensambladas mediante FLASH se unirán en vectores de expresión TALEN mediante el uso de 400 U de ADN ligasa T4 (New England Biolabs). Los productos de unión se transformarán en células XL-1 Blue químicamente competentes. Típicamente, se escogerán seis colonias para cada unión y se aislará el ADN plasmídico mediante un procedimiento Miniprep de lisis alcalina. Simultáneamente, las mismas colonias se examinarán por PCR mediante el uso de los cebadores oSQT34 (5'-GACGGTGGCTGTCAAATACCAAGATATG-3') y oSQT35 (5'-TCTCCTCCAGTTCACTTTTGACTAGTTGGG-3'). Los productos de PCR se analizarán en un sistema de electroforesis capilar QIAxcel (Qiagen). El ADN miniprep de los clones que contenían productos de PCR de tamaño correcto se enviará para la confirmación de la secuencia de ADN con los cebadores oSQT1 (5'-AGTAACAGCGGTAGAGGCAG-3'), oSQT3 (5'-ATTGGGCTACGATGGACTCC-3'), y oJS2980 (5'-TTAATTCAATATATTCATGAGGCAC-3'); oSQT1 hibrida en el extremo 5' de la secuencia codificante del conjunto de repeticiones TALE y permitirá la secuenciación de la mitad amino-terminal del conjunto ensamblado, oSQT3 hibrida en el extremo 3' de la secuencia codificante del conjunto de repeticiones TALE y permite la secuenciación de la mitad carboxi terminal del conjunto ensamblado, y oJS2980 se prepara dentro de la secuencia codificante del dominio FokI (aguas abajo de oSQT3) y permitirá la secuenciación y verificación del dominio de repetición de TALE 0,5 carboxi terminal.

Tamizaremos seis colonias para cada ensamblaje como se describió anteriormente, seguido de seis colonias adicionales si es necesario. Con este enfoque, uno o más clones verificados en secuencia para >90 % de las reacciones de ensamblaje. Estos porcentajes se derivarán principalmente de experimentos diseñados para construir fragmentos de ADN que codifican 16,5 repeticiones TALE.

Actividad TALEN EGFP y ensayos de toxicidad

Los ensayos con indicador EGFP se realizarán en una línea celular humana U2OS clonal que porta un constructo integrado que expresa constitutivamente una proteína de fusión EGFP-PEST. Esta línea clonal se derivará de una línea indicadora policlonal EGFP-PEST U2OS. Las células EGFP-PEST U2OS clonales se cultivarán en DMEM avanzado (Life Technologies) suplementado con FBS al 10 %, GlutaMax (Life Technologies) 2 mM, penicilina/estreptomycin y 400 μ g/ml de G418. Las células se transfectarán por triplicado con 500 ng de cada ADN plasmídico TALEN y 50 ng de ADN plasmídico ptdTomato-N1 mediante el uso de un sistema 4D-Nucleofector Lonza, solución SE y el programa DN-100 de acuerdo con las instrucciones del fabricante. Como control negativo se transfectará por triplicado solamente 1 μ g de plásmido ptdTomato-N1. Las células se analizarán para determinar la expresión de EGFP y tdTomato a los 2 y 5 días después de la transfección mediante el uso de un citómetro de flujo BD FACSAriaII.

Amplificación por PCR y verificación de secuencia de genes humanos endógenos

Las reacciones de PCR para amplificar los loci objetivos se realizarán mediante el uso de los cebadores que se muestran en la Tabla complementaria 5. Se realizarán condiciones de PCR estándar con ADN polimerasa de alta fidelidad Phusion Hot Start II (Thermo-Fisher) de acuerdo con las instrucciones del fabricante durante 35 ciclos (98 °C, desnaturalización

10 s; 68 °C, hibridación 15 s; 72 °C, extensión 30 s). Para los loci que no se amplifican en condiciones estándar, usaremos una de las siguientes modificaciones: 1) la adición de betaína a una concentración final de 1,8 M, 2) PCR touchdown ([98 °C, 10 s; 72-62 °C, -1 °C/ciclo, 15s; 72 °C, 30s]_{10 ciclos}, [98 °C, 10 s; 62 °C, -1 °C/ciclo, 15s; 72 °C, 30 s]_{25 ciclos}) con betaína 1,8 M, y 3) la adición de DMSO al 3 % o 5 % y una temperatura de hibridación de 65 °C. Los productos de PCR se analizarán para determinar el tamaño correcto en un sistema de electroforesis capilar QIAxcel. Los productos de tamaño correcto se tratarán con ExoSap-IT (Affymetrix) para eliminar los nucleótidos o cebadores no incorporados y se enviarán para secuenciar el ADN para confirmar la secuencia del gen endógeno.

Ensayo de endonucleasa I T7 para cuantificar la mutación mediada por NHEJ de genes humanos endógenos

Las células U2OS-EGFP se cultivarán y transfectarán por duplicado como se describió anteriormente. El ADN genómico se aisló a partir de células transfectadas con plásmidos control o codificadores de TALEN mediante el uso de un sistema de purificación basado en perlas magnéticas de alto rendimiento (Agencourt DNAdvance) de acuerdo con las instrucciones del fabricante. La PCR para amplificar loci endógenos se realizará durante 35 ciclos como se describió anteriormente y los fragmentos se purificarán con Ampure XP (Agencourt) de acuerdo con las instrucciones del fabricante. 200 ng de producto de PCR purificado se desnaturalizarán y se volverán a hibridar en NEBuffer 2 (New England Biolabs) mediante el uso de un termociclador con el siguiente protocolo (95 °C, 5 min; 95-85 °C a -2 °C/s; 85-25 °C a -0.1 °C/s; mantener a 4 °C).³³ Los productos de PCR hibridados se trataron con 10 U de Endonucleasa I T7 a 37 °C durante 15 minutos en un volumen de reacción de 20 µl. Las reacciones se detuvieron mediante la adición de 2 µl de EDTA 0,5 M, se purificaron con Ampure XP y se cuantificaron en un sistema de electroforesis capilar QIAxcel mediante el uso del método OM500. La suma del área debajo de los picos de escisión específicos de TALEN (expresada como un porcentaje del pico del amplicón parental, indica una fracción escindida) se usa para estimar los niveles de modificación génica mediante el uso de la siguiente ecuación como se describió previamente.

$$(\% \text{ de modificación de gen} = 100 \times (1 - (1 - \text{fracción escindida})^{1/2})$$

Ejemplo 2

Se sintetizaron cinco fragmentos que se muestran más abajo y cada uno se clonó en un pUC57 modificado: pUC57-ΔBsaI (vectores como se describe en Juong y otros artículos sobre ensamblaje mediante FLASH). Contiene un solo cambio de pares de bases para romper un sitio de BsaI con XbaI y BamHI.

EBE de RTN1:

NK:

XbaI BbsI

ATGCA T[^]CTAGA-GAAGACAA[^]CTGA-

GCACCGAGCAGGTGGTGGCCATCGCCAGCAACAAGGGCGGCAAGCAGGCCCTG

GAGGCCGTGAAGGCCACCTGCTGGACCTGCTGGGCGCCCCCTACGAG-

CTGA[^]AGAGACC-G[^]GATCC(CGGGC)

BsaI

BamHI

NN:

ATGCA

TCTAGAGAAGACAAGTCTGAGCACCGAGCAGGTGGTGGCCATCGCCAGCAACAACG

GCGGCAAGCAGGCCCTGGAGGCCGTGAAGGCCAGCTGCTGGAGCTGAGGGCCG

CCCCCTACGAGCTGAAGAGACCGGATCC CGGGC

NG:

ATGCA

TCTAGAGAAGACAAGTCTGAGCACCG^{gag}CAGGTGGTGGCCATCGCCAGCAACGGCG

GCGGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACC

GGCCCCCTACGAGCTGAAGAGACCGGATCC CGGGC

HD:

ATGCA

TCTAGAGAAGACAACCTGAGCACCGagCAGGTGGTGGCCATCGCCAGCCACGACG
GCGGCAAGCCCCGCCCTGGAGGCCGTGTGGGCCAAGCTGCCCCGTGCTGAGGGGCG
TGCCCTACGAGCTGAAGAGACCGGATCC CGGGC

Sl:

ATGCA

TCTAGAGAAGACAACCTGAGCACCGAGCAGGTGGTGGCCATCGCCAGCAGCATCG
GCGGCAAGCAGGCCCTGGAGGCCGTGAAGGTGCAGCTGCCCCGTGCTGAGGGCCG
CCCCCTACGAGCTGAAGAGACCGGATCC CGGGC

Como prueba de principio, estos fragmentos clonados se usarán para generar proteínas quiméricas de seis unidades de repetición fusionadas a la nucleasa FokI, mediante el protocolo exacto en el artículo sobre las TALEN mediante FLASH de Joung, es decir, una proteína quimérica que se dirige a una cadena de nucleótidos A (C, T y G). A continuación, estas proteínas quiméricas se evaluarán para determinar la eficiencia de unión/direccionamiento a las bases de ADN deseadas mediante el uso de un constructo indicador.

Una vez que se confirma la eficiencia de unión de estas unidades, se generará una biblioteca de EBE de *Ralstonia* que será una copia exacta de la biblioteca de EBE de *Xanthomonas* de TALEN mediante FLASH. Esta biblioteca puede usarse para generar TALEN de *Ralstonia* mediante el protocolo exacto del sistema de TALEN mediante FLASH.

Ejemplo 3. Generación de vectores de ácido nucleico con metilesterasa

Las secuencias adicionales clonadas a partir de otras especies o clonadas a partir de la misma especie pueden usarse funcionalmente como una enzima ya sea por sí mismas o en serie como un monómero o polímero (fusión de proteínas) para realizar cualquiera de los experimentos descritos en la presente descripción con reconocimiento de ADN. Se creó una secuencia consenso de identificación de RVD mediante el uso de técnicas de optimización de secuencia conocidas en la técnica. Se realizó una búsqueda BLAST a lo largo de secuencias de metilesterasa en especies bacterianas (ver Figura 1). Se identificaron los siguientes polipéptidos con capacidad de reconocimiento de pares de bases de ADN similares a las secuencias de ácido nucleico y los polipéptidos descritos en la presente SEQ ID NO:1-19:

EBE de TAL contra metilesterasas

#1
EBE de consenso de <i>Xanthomonas</i>
LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG
LTPEQVVAIANNNGGKQALETVQRLLPVLCQAHG
LTPDQVVAIASHDGGKQALETVQRLLPVLCQAHG
LTPAQVVAIASNIGGKQALETVQRLLPVLCQDHG
<u>EJO92907</u> 166 TTDRVVALGTSTGGTQALEVVLRLPVDC 194
<u>YP_001187060</u> 166 TTDRVVALGTSTGGTQALEVVLRLPVDC 194
<u>YP_003847734</u> 169 MTSEQIVAIGTSTGGTQALEAVLTALPRVC 198
<u>ZP_08780698</u> 167 TTDRVVAIGTSTGGTQALEVLTALPRVC 195
<u>YP_004846745</u> 185 TTERIVAIGTSTGGTQALETVLHRLPATC 213
<u>YP_005027668</u> 186 TTDKIIAIGTSTGGTQALEAVLTKLPAVC 214
<u>ZP_10991552</u> 174 TTERIVAIGTSTGGTQALETVLTALPRVC 202
<u>YP_001792820</u> 162 TTERVVALGTSTGGTQALEVVLRLTPRVC 190
<u>EKE17764</u> 172 TTDQLIAIGTSTGGTQALEAILTKLPAVC 200
<u>ZP_03698248</u> 178 TTERIVAIGTSTGGTQALETVLPRLPATC 206
<u>EGH48032</u> 11 TTERIVAIGTSTGGTQALEAVLTALPRVC 39

ES 2 778 033 T3

	<u>ZP_06495900</u> 1 TTERIVAIGTSTGGTQALEAVLTALPRVC 29
	<u>ZP_10381001</u> 76 TTERIVAIGTSTGGTQALEAVLTALPRVC 104
5	<u>ZP_10442431</u> 158 TSDKVVAIGASTGGTQALELLLTGLPAVC 186
	#2
10	<u>ZP_10991552</u> 174 TTERIVAIGTSTGGTQALETVLTALPRVC 202
	<u>EGH48032</u> 11 TTERIVAIGTSTGGTQALEAVLTALPRVC 39
	<u>ZP_06495900</u> 1 TTERIVAIGTSTGGTQALEAVLTALPRVC 29
15	<u>EGH61007</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>EGH06695</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>EGH31878</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
20	<u>EGH66597</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>ZP_07003572</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>ZP_06457223</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
25	<u>ZP_04590480</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>ZP_07251539</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>NP_790747</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>EGH77388</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
30	<u>EFW86187</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>EGH54563</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>YP_233877</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
35	<u>EGH23390</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>ZP_05638023</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>EGH71924</u> 106 TTERIVAIGTSTGGTQALEAVLTALPRVC 134
40	<u>EFW82095</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>ZP_07265841</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
	<u>YP_273082</u> 172 TTERIVAIGTSTGGTQALEAVLTALPRVC 200
45	<u>YP_004030667</u> 117 FSQADIVRIADNIGGAQALKAVLEHGPTL 145
	<u>YP_004030667</u> 186 ADIVKIASNNGGGAQALEAVAMHGSTLCE 213
	<u>YP_004030667</u> 153 ADIVKIAGNNGGARALKAVVMHGPTLCE 180
50	<u>ZP_10995147</u> 155 TTDRVVALGCSTGGTQALEFILRQLPRDC 183
	<u>EGH56182</u> 30 ALAAVGGKGALEVPANLIPANCE 53
	<u>YP_003907367</u> 173 RIVAIGTSTGGTQALEVVLTALP 195
55	EBE1 LTPDQVVAIASNNGGKQALETVQRLLPVLCQDHG 34
	EBE4 LTPAQVVAIASNNGGKQALETVQRLLPVLCQDHG 34
	EBE3 LTPDQVVAIASHDGGKQALETVQRLLPVLCQAHG 34
	EBE2 LTPEQVVAIANNNNGGKQALETVQRLLPVLCQAHG 34
60	<u>ZP_07265841_2</u> -TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	<u>YP_273082_2</u> -TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	<u>EFW82095_2</u> -TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
65	<u>EGH71924_2</u> -TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29

ES 2 778 033 T3

	ZP_05638023_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EGH23390_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
5	YP_233877_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EGH54563_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EFW86187_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
10	EGH77388_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	NP_790747_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	ZP_07251539_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
15	ZP_04590480_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	ZP_06457223_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	ZP_07003572_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
20	EGH66597_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EGH31878_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EGH06695_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
25	EGH61007_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	ZP_06495900_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EGH48032_2	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
30	ZP_10381001	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	ZP_06495900	-TTERIVAIGTSTGGTQALEAVLTALPRVC---- 29
	EGH48032	-TTERIV AIGTSTGGTQALEA VLTALPRVC---- 29
35	YP_003847734	MTSEQIVAIGTSTGGTQALEAVLTALPRVC---- 30
	ZP_10991552	-TTERIVAIGTSTGGTQALETVLTALPRVC---- 29
	ZP_10991552_2	-TTERIVAIGTSTGGTQALETVLTALPRVC---- 29
	YP_003907367_2	----RIVAIGTSTGGTQALEVVLTALP----- 23
40	EJO92907	-TTDRVVALGTSTGGTQALEVVLRQLPVDC---- 29
	YP_001187060	-TTDRVVALGTSTGGTQALEVVLRQLPVDC---- 29
	ZP_10995147_2	-TTDRVVALGCSTGGTQALEFILRQLPRDC---- 29
45	YP_001792820	-TTERVVALGTSTGGTQALEVVLRTLPRVC---- 29
	ZP_08780698	-TTDRVVAIGTSTGGTQALEVVLTLALPRVC---- 29
	YP_004846745	-TTERIVAIGTSTGGTQALETVLHRLPATC---- 29
50	ZP_03698248	-TTERIVAIGTSTGGTQALETVLPRLPATC---- 29
	YP_005027668	-TTDKIIAIGTSTGGTQALEAVLTKLPAVC---- 29
	EKE17764	-TTDQLIAIGTSTGGTQALEAILTKLPATC---- 29
55	ZP_10442431	-TSDKVVAIGASTGGTQALELLLTGLPAVC---- 29
	YP_004030667_2b	---ADIVKIASNGGGAQALEAVAMHGSTLCE--- 28
	YP_004030667_2c	---ADIVKIAGNGGGARALKAVVMHGPTLCE--- 28
	YP_004030667_2a	FSQADIVRIADNIGGAQALKAVLEHGPTL----- 29
60	EGH56182_2	-----ALAAAVGGKGALEVPANLIPANCE--- 24

Ejemplo 4

65 Un par de EBE específicos de Bmpr2 (dominio de unión al ADN de Ralstonia, 16EBE cada uno) se sintetizaron por genes y se clonaron en XTN-BB (cadena principal de TAL de Xanthomonas fusionada a FokI). Estos constructos se

cotransfectaron en células C6 de rata y se extrajo el ADNg después de 48 horas para el ensayo de nucleasa Cell de Surveyor. Si tiene éxito, el ensayo debe producir subpoblaciones de 240 pb y 150 pb a partir del amplicón original de 400 pb del locus. Los resultados se muestran en las Figuras 2.

El ensayo revela las bandas esperadas de 250 pb y 150 pb en las células transfectadas con TALEN de *Ralstonia* y *Xanthomonas*, que están ausentes en el control negativo WT. Esto indica que los EBE de *Ralstonia* se dirigen a este locus y que la fusión de la nucleasa FokI a los EBE de *Ralstonia* conduce a una digestión dirigida del ADN genómico. Mediante el uso de la banda de 250 pb, 5,75 % para XTN, 1,82 % para RTN. Mediante el uso de la banda de 150 pb, 3,66 % para XTN, 5,43 % para RTN.

Sitio objetivo de Bmpr2	T-T-GATA-GTCG-CCTT-ATG-T tttgatacagaatgt T-GAC-AGGT-AAAC-GAAA-T-A
RTN directo	TGATAGTCGCCTTATG
RTN inverso	ATTTGGTTTACCTGTC
Nota: los RTN no especifican el primer y el último nucleótido del sitio objetivo (subrayado). Estos se especifican por la cadena principal de TALEN de <i>Xanthomonas</i> .	

Secuencia de aminoácidos de EBE de RTN directa de Bmpr2

LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG

LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV
 LSTEQVVAIAS NN GGKQALEAVKAQLELRAAPYE
 LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
 LSTEQVVAIAS NN GGKQALEAVKAQLELRAAPYE

LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV
 LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
 LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPYA
 LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV

LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPYA
 LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPYA
 LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
 LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG

LSTEQVVAIAS NN GGKQALEAVKAQLELRAAPYE
 LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
 LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV

Secuencia de ADN de RTN directa de Bmpr2
 (fuente en negritas: EBE de *Ralstonia* sintetizados) esta secuencia es contigua:

5 GACGGATCGGGAGATCTCCCGATCCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTG
 CTCCCTGCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATT
 GCATGAAGAATCTGCTTAGGGTTAGGCGTTTTCGCGCTGCTTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTG
 ACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
 TGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG
 ACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
 10 GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
 AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGAC
 TCACGGGGATTTCGAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCAAAAATG
 TCGTAACAACCTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTA
 ACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCACCATTGGAC
 TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCAAGAAGAAG
 AGGAAGGTGGGCATTCACGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTTCGCAACAGCAACAGGAGAAAATC
 AAGCCTAAGGTCAGGAGCACCGTCGCGCAACACACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGC
 TTTCACAGCACCCCTCGGCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGA
 15 GGCAATTGTAGGGGTTCGTTAAACAGTGGTCCGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGG
 GCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGC
 CTGGCGCAATGCGCTCACCGGGGCCCCCTTGAAC
 CTGAGCACCGCCCAGGTGGTGGCCATCGCCAGC AACGGC
 GGCGGCAAGCAGGCCCTGGAGGGCATCGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
 20 CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAAG
 GGCGGCAAGCAGGCCCTGGAGGGCGTGAAGGCCACCTGCTGGACCTGCTGGGCGCCCCCTACGTG
 CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAAC
 GGCGGCAAGCAGGCCCTGGAGGGCGTGAAGGCCAGCTGCTGGAGCTGAGGGCGCCCCCTACGAG
 CTGAGCACCGCCCAGGTGGTGGCCATCGCCAGC AACGGC
 GGCGGCAAGCAGGCCCTGGAGGGCATCGCGGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
 CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAAC
 GGCGGCAAGCAGGCCCTGGAGGGCGTGAAGGCCAGCTGCTGGAGCTGAGGGCGCCCCCTACGAG
 25 CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAAG
 GGCGGCAAGCAGGCCCTGGAGGGCGTGAAGGCCACCTGCTGGACCTGCTGGGCGCCCCCTACGTG
 CTGAGCACCGCCCAGGTGGTGGCCATCGCCAGC AACGGC
 GGCGGCAAGCAGGCCCTGGAGGGCATCGCGGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
 CTGAGCACCGCCCAGGTGGTGGCCATCGCCAGC CACGAC
 GGCGGCAAGCCCCCTGGAGGGCGTGTGGGCCAAGCTGCCCGTCTGAGGGCGCTGCCCTACGCC
 30 CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAAG

GGCGGCAAGCAGGCCCTGGAGGCCGTGAAGGCCACCTGCTGGACCTGCTGGGCGCCCCCTACGTG

CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC CACGAC
GGCGGCAAGCCCGCCCTGGAGGCCGTGGGCCAAGCTGCCCGTCTGAGGGCGTGCCCTACGCC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC CACGAC
GGCGGCAAGCCCGCCCTGGAGGCCGTGGGCCAAGCTGCCCGTCTGAGGGCGTGCCCTACGCC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC
GGCGGCAAGCAGGCCCTGGAGGCCATCGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC
GGCGGCAAGCAGGCCCTGGAGGCCATCGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC

CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAA
GGCGGCAAGCAGGCCCTGGAGGCCGTGAAGGCCAGCTGCTGGAGCTGAGGGCGCCCCCTACGAG
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC
GGCGGCAAGCAGGCCCTGGAGGCCATCGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAA
GGCGGCAAGCAGGCCCTGGAGGCCGTGAAGGCCACCTGCTGGAGCTGCTGGGCGCCCCCTACGTG

CTGAGCACCGCCAGGTGGTGGCCATCGCCAGCAACGGCGAGGACGGCCAGCCTTGGAGTCCATCTGAGCCCAATTGTCCA
GGCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGC
AGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTTGATCAAAAGAACCAACCGGCGGATCCCGAGAGAACTTCCCATCGAGTC
GCGGGATCCCAACTAGTCAAAAGTGAAGTGAAGGAGAAATCGAACTTCGTCATAAATGAAATATGTGCCTCATGAAT
ATATTGAATTAATTGAAATTTGCCAGAAATTCCTACTCAGGATAGAATTTCTGAAATGAAGGTAATGGAATTTTTTATGAAAGTT
TATGGATATAGAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCGGATCTCCTATTGATTACG
GTGTGATCGTGGATCAAAAGCTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACTTAGAGGAAAGTCAGACGGAAATTT
AGAAAAATCAACACGAAACAAACATATCAACCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAAATTTAAGTTT
TTATTGTGAGTGGTCACTTTAAAGGAACTACAAAGCTCAGCTTACACGATTAATCATATCACTAATTGTAATGGAGCTGT
TCTTAGTGTAGAAAGAGCTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACTTAGAGGAAAGTCAGACGGAAATTT
AATAACGGCGAGATAAACTTTAAGGGCCCTTCGAAGGTAAGCCTATCCCTAACCTCTCCTCGGTCTCGATTCTACGCTAC
CGGTCACTACACCACTACCATTTGAGTTTAAACCCGCTGATCAGCTTCGACCTTCGACCTTCGCTTCCTAGTTTGGTGGT
CCCCCCCCCGTGGCTTCTTACCCCTGGAAGGTGCCACTCCCACTGCTCTTCTTCTAATAAATGAGGAAATTTGATCGCAT
GTCTGAGTAGGTGTCTATTCTTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAAATGACAGGC
ATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACGGCCCTG
TAGCGGGCAATTAAGCGCGGGGGTGTGGTGGTTACGGCAGCGTGACCGCTACACTTGGCAGCGCCCTAGCGCCCGCTCT
TTCGCTTCTTCTCCCTTCTTCTTCTCGCCACGTTCCGCGGCTTTCGCCGTCAGCTCTAAATCGGGGCATCCCTTTAGGGTTCGA
TTTAGTGTCTTACGGCACCTCGACCCCAAAAACCTTGAATAGGGTGTGGTTCACGTAAGTGGGCCATCGCCCTGATAGACGG
TTTTTCGCCCTTTGACGTGTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCAAAACCTGGAACAACACTCAACCTATCTCGG
TCTATTCTTTGATTTATAAGGGATTTTGGGGATTTTCGGCTATTGGTTAAAAAATGAGCTGATTTAAACAAAAATTTAACGCG
AATTAATCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAAGTATGCAAAAGCATGC
ATCTCAATTAGTACAGCAACCAAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAAGTATGCAAAAGCATGCATCTCAATTA
GTACAGCAACCATAGTCCCGCCCTAACTCCGCCCATCCGCCCTTAACCTCCGCCAGTTCCGCCCATCTTCCGCCCATGGCT
GACTAATTTTTTTTATTTATGACAGAGCCGAGGCCGCTCTGCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGG
GGCTAGGCTTTTGCAAAAAGTCCCGGAGCTTGTATATCTTTTCGGATCTGATCAGCAGCTGTGTGACAATTAATCATCTC
GCATAGTATATCGGCATAGTATAATACGCAAGGTGAGGAATTAACCATGGCCAAGCCTTTGTCTCAAGAAGAATCCACCC
TCATTGAAGAGCAACGGCTACAATCAACAGCATCCCTACTCTGAAGACTACAGCTCGCCAGCGCAGCTCTCTTCTAGCGA
CGCGCGCATCTTCACTGGTGTCAATGTATATCATTTTACTGGGGACCTTGTGCAGAACTCGTGGTGTCTGGGCACCTGCTGCTG
CTGCGGACGTGGCAACCTGATCTGTATCTGTCGCGATCGGAAATGAGAACAGGGGCATCTTGAAGCCCTGCGGACGGGTGTCG
ACAGTGTCTTCTCGATCTGATCTTGGGATCAAAAGCATAGTGAAGGACAGTGTGAGCAGCCGACGGCAGTTGGGATCTCTG
GAATTGCTGCCCTCTGGTTATGTGTGGGAGGGCTAAGCACTTCTGTTGGCGAGGAGCAGGACTGACACGTGCTACGAGATTTC
GATCTCACCGCCGCTTCTATGAAGGTTGGGCTTCGGAAATCGTTTTCGGGACGCCGGCTGGATGATCTCTCCAGCGCGGG
ATCTCATGTGGAGTCTTTCGCCCAACCCCACTTGTATTATGACGCTATAATGGTTACAATAAAGCAATAGCATCACAAAT
TTCACAAATAAAGCATTTTTTCTACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGTATACCG
TCGACCTCTAGCTAGAGCTTGGCGTAATCATGGTTCATAGCTGTTCCTGTGTGAAATTTGTATCCGCTCACAAATTCACACAA
CATACGAGCCGGAAGCATAGTAAAGCTTAAAGCCTGGGGTGCTTAATGAGTGAAGTAACTACATCAATTAATTGCGTTGGCTCACTG
CCCCGCTTTCAGTCTGGGAAACCTGTCTGTGCCAGTGCATTAATGAATCGGCCAACCGCGCGGGGAGAGGCGGTTTTCGCTATTG
GGCGCTCTTCGCTTCTCTGCTACTGACTCGCTGCGCTCGGTCTCGGTGCGGCGAGCGGTATCAGCTCACTCAAAAGGCG
GTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGGAACCG
TAAAAAGGCCGCTTGTGTGGCTTTTTCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAGAGG
TGGCGAAACCCGACAGGACTATAAAGATACCAAGCGTTTCCCTGGAAGCTCCCTCGTGGCTCTCTGTTCGACCCCTGCC
GCTTACCGGATACCTGTCCGCTTCTTCCCTTCGGGAAGCGTGGCGCTTCTCAATGCTACGCTGTAGGTATCTCATGTTGCGT
GTAGGTTCGTTCCGCTCAAGCTTGGGCTGTGTGACGAACCCCGCTTACGCCCCGACCGCTTACGCTTATCCGGTAACTATCTGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCTGGTAACAGGATTAGCAGAGCGAGGTATGTAG
GCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAA
GCCAGTTACCTTCGGAAGAAAGAGTTGGTAGCTCTGATCGCGCAAAACAAACCACCGCTGTTAGCGGTGGTTTGTGTGCA
AGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCTTTGATCTTTTACGGGGTCTGACGCTCAGTGGAAACGA
AAACTACGTTAAGGGATTTTGGTATGAGATTATCAAAAAGGATCTTCACTAGATCTTTTAAATTAATAAATGAAGTTTAA
AATCAATCTAAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGT
CTATTTCGTTATCCATAGTTGCTGACTCCCGCTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
TGCAATGATACCGCGAGACCCACGCTACCGGCTCCAGATTATACGCAATAAACACGACCCGGAAGGGCCGAGCGCAG
AAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTGCGCAGTTAATA
GTTTGGCGAACGTTGTTGCCATTGCTACAGGCACTCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTAGCTCCGGTTCCC
AACGATAAGCGGAGTTACATAGCAAGTTCCTTGTGTGCAAAAAGCGGTTAGCTCTTCCGCTCCGATCGTTGTGAGAAG
TAAGTTGGCCCGAGTGTATCACTCATGGTATGGCAGCACTGCATAATTTCTTACTGTCAATGCCATCCGTAAAGATGCTTTTC
TGTGACTGGTGAATCAACCAAGTCTTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGGCCGCGCTCAATACGG
GATAATACCGCGCCACATAGCAAGATTTAAAGGTGCTCATGATGGAAACGTTCTTCCGCGCGGAAACCTCAAGGATCT
TACCGCTGTGTGAGTCCAGTTGATGTAAACCACTCTGTCAGCCCACTGATCTTACGATCTTTTACTTTCACCGCGTTTCTG
GGTGAGCAAAAAAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATCTCATACTCTTCC
TTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTGAATGTATTGAAAAAATAAACAA
ATAGGGGTTCGCGCACATTTCCCGGAAAAGTGCCACCTGACGTC

Secuencia de aminoácidos de EBE de RTN inversa de Bmpr2

5 LSTEQVVAIAS NN GGKQALEAVKAQLLELRAAPYE
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPIA
10 LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
15 LSTEQVVAIAS NN GGKQALEAVKAQLLELRAAPYE
LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPIA
LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPIA
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
20 LSTEQVVAIAS NK GGKQALEAVKAHLLDLLGAPYV
LSTAQVVAIAS NG GGKQALEGIGEQLLKLRTAPYG
LSTAQVVAIAS HD GGKPALEAVWAKLPVLRGVPIA
25

Secuencia de ADN de RTN inversa de Bmpr2

(fuente en negritas: EBE de Ralstonia sintetizados) esta secuencia es
contigua:

30 GACGGATCGGGAGATCTCCCGATCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATCTG
CTCCCTGCTTGTGTGTGGAGGTGCTGAGTAGTGCGGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAATT
GCATGAAGAATCTGCTTAGGGTTAGGCGTTTTCGCGTGTCTCGCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTG
ACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
35 TGGCCCGCTGGCTGACCGCCCAACGACCCCGCCCATGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG
ACTTTCCATTGACGTCAATGGGTGGA^aTATTACGGTAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGAC
TCACGGGGATTTC^aAGTCTCCACCCCATGACGTCAATGGGAGTTTGT^aTTTGGCACCAAAATCAACGGGACTTTC^aAAAATG
40 TCGTAACA^aACTCCGCCCCATTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTA
ACTAGAGA^aCCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCACCATGGAC
TACAAAGACCATGACGGTGTATTATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGAAGAAG
AGGAAGGTGGGCATTACCGCGGGGTACCTATGGTGGACTTGAGGACACTCGGTTATTTCGCAACAGCAACAGGAGAAAAATC
AAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTCGCGC
45 TTTCACAGCACCCCTGCGGGCGTTGGGACGGTGGCTGTCAAAATACCAAGATATGATTGCGGGCCCTGCCCGAAGCCACGCACGA
GGCAATTGTAGGGGTGCGTAAACAGTGGTGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTAGGGG
GCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTGCACGC
CTGGCGCAATGCGCTACCGGGGCCCCCT
45 CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAA^a GGC^aGCAAGCAGGCCCTGGAGGCCGTGAAGGCCAGCTGCTGGAGCTGAGGGCCGCCCTACGAG
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC GGC^aGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC GGC^aGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC GGC^aGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
50 CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACAA^a GGC^aGCAAGCAGGCCCTGGAGGCCGTGAAGGCCACCTGCTGGACCTGCTGGGCGCCCCCTACGTG
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC GGC^aGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC GGC^aGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCCCTACGGC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACAA^a GGC^aGCAAGCAGGCCCTGGAGGCCGTGAAGGCCAGCTGCTGGAGCTGAGGGCCGCCCTACGAG
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC CACGAC GGC^aGCAAGCCGCCCTGGAGGCCGTGGGGCAAGCTGCCGTGCTGAGGGCGTGCCTACGCC
55 CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC CACGAC GGC^aGCAAGCCGCCCTGGAGGCCGTGGGGCAAGCTGCCGTGCTGAGGGCGTGCCTACGCC

CTGAGCACCCGACGGTGGTGGCCATCGCCAGC AACGGC GGC GGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCTACGGC
CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC AACAAAG GGC GGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCTACGGT
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC AACGGC GGC GGCAAGCAGGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCTACGGC
CTGAGCACCGCCAGGTGGTGGCCATCGCCAGC CACGAC GGC GGCAAGCCGCCCTGGAGGGCATCGGCGAGCAGCTGCTGAAGCTGAGGACCGCCCTACGGC
CTGAGCACCGAGCAGGTGGTGGCCATCGCCAGC
5 AGCATCGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCAAATTGTCCAGGCCGATCCCGCGTTGGCTGCGTTAACGAATG
ACCATCTGGTGGCGTTGGCATGTCTTGGTGGACGACCCGCGCTCGATGCAGTCAAAAAAGGGTCTGCGCTCATGCTCCCGCATT
GATCAAAAAGAACC AACCGGCGGATTCGCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAACTAGTCAAAAAGTGAACCTGGA
GGAGAAGAAATCTGAACCTTCGTCATAAAATTGAAATATGTGGCTCATGAATATATTGAATTAATTGAAATTGCCAGAAATTCC
ACTCAGGATAGAATTCTTGAATGAAGGTAATGGAAATTTTTATGAAAGTTTATGGATATAGAGGTAAACATTTGGGTGGAT
10 CAAGGAAACCGGACGGAGCAATTTATACGTGCGGATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGG
AGGTTATAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTGCAAGAAAAATCAAAACAGAAACAAACATATCAA
CCCTAATGAATGGTGGAAAGTCTATCCATCTTCTGTAACGGAAATTTAAGTTTATTTGTGAGTGGTCACTTTAAAGGAAACT
ACAAAGCTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGTAGAAGAGCTTTTAATTGGTGGAA
GAAATGATTAAAGCCGGCACATTAACCTTAGAGGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTTAAGGGCCC
15 TCGAAGGTAAGCTATCCCTAACCTCTCGGTCTCGATTCTACGCGTACCGGTCAATCATCACTACCATCTCCATTTAGTTTA
AACCCGCTGATCAGCTCGACTGTGCCTTCTAGTTGCCAGGCATCTGTTGTTTGCCCTCCCGCGTGCCTTCTTGACCCTGGA
AGGTGCCACTCCCAGTGTCTTCTTAATAAAAAATGAGGAAATGTCATCGCATTTGTCTGAGTAGGTGTCACTTCTTCTGGGG
GTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCATATGGCTT
15 CTGAGGCGGAAAGAACAGCTGCGGCTCTAGGGGGTATCCCCACGCGCCCTGTAGCGGCGCATTAAGCGCGCGGGGTGG
TGGTTACGCGCAGCGTGACCGCTACACTTGGCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTCTTCCCTTCTTCTCGCCACGT
TCGCCGGCTTTCGCCGCTAAGCTCTAAATCGGGGCATCCCTTTAGGGTTCCGATTAGTGCTTTACGGCACCTTAAAGCCAA
AACTTGATTAGGGTGTGTTTACGTAGTGGGCCATCGCCCTGATAGACGGTTTTCGCGCTTTGACGTTGGAGTCCACGTT
20 CTTTAATAGTGGACTTGTTCACAACTGGAACAACACTCAACCTATCTCGGTCTATTCTTTGATTATAGAGGATTTTGGG
GATTTTCGGCTTATTGGTTAAAAAATGAGCTGATTTAAACAAAAATTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGG
GTGTGGAAGTCCCCAGGCTCCCCAGGCAGGCAAGATATGCAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAA
AGTCCCCAGGCTCCCCAGCAGGCAAGATATGCAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAACTCC
25 GCCATCCCGCCCTAACTCCGCCAGTTCGCCCTTCTCCGCCCATGGCTGACTAATTTTTTTTTTTTATGTCAGAGGCCGA
GGCCGCTCTGCTCTGAGCTATTCAGAAAGTAGTAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGA
GCTTGATATCCATTTTCGGATCTGATCAGCAGCTGTGTGACAATTAATCATCGGCATAGTATATCGGCATAGTATAATACGAC
AAGGTGAGGAACATAACCATGGCCAAGCCTTTGTCTCAAGAAAGAAATCCACCCTCATTGAAAGAGCAACGGCTACAATCA
25 AGCATCCCCATCTCTGAAGACTACAGCGTCGCCAGCGCAGCTCTCTAGCGACGGCCGCATCTTCACTGGTGTCAATGTATA
TCTTTTACCTGGGGAGCTTGTGCAAGACTCGTGGTGTGCTGGGCACTGCTGCTGCTGCGGCAGCTGGCAACCTGACTTGTATCG
TCGCGTAGGAAATGAGAACAGGGGCTTTGAGCCCTTGGCGGCTGCGGAGGTGTCGACAGGTGCTTCTGATCTGCACTGGGAT
CAAAGCGATAGTGAAGGACAGTGATGGACAGCCGACGGCAGTTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTGGGAG
GGCTAAGCACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATTTCGATTCCACCGCCGCTTCTATGAAAGGTTG
30 GGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCTCCAGCGCGGGGATCTCATGTGGAGTTCTTCGCCACCCCAA
CTTGTTTATGTCAGCTTATAATGGTTACAAATAAAGCAATAGCATCAAAAATTTACAAATAAAGCATTTTTCATCTGCAAT
CTAGTTGTGTTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGTAATCAT
GGTCATAGCTGTTTTCTGTGTGAATTTGTTATCCGCTCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC
35 TCGGGGTGCCAATAGTAGAGTAACTACATTAATTTGCGTTGCGCTCACTGCCCCGTTTCCAGTCGGGAAACCTGTGCTGGC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGGCTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTC
GCTGCGCTCGGTCTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGTTATCCACAGAATCAGGGGAT
40 AACGCAGGAAAGAACATGTGAGCAAAAAGGCCAGCAAAAAGGCCAGGAACCGTAATAAAGGCCGCGTTGCTGGCGTTTTTCAT
AGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAAGAGGTGGCGAAACCCGACAGGACTATAAAGATAC
CAGGCGTTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCGGCTTACCGGATACCTGTCCGCTTTCTCCCT
TCGGGAAGCGTGGCGCTTTCTCAATGCTCAAGCTGTAGGTATCTCAGTTCCGTTGAGGTGCTTCGCTCAAGCTGGGCTGTGT
40 GCACGAACCCCGGTTACGCGGACCGCTGCGCCTTATCCGGTAACATCGCTTGTAGTCCAACCCGTAAGACACGACTTAT
CGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGCGGTGCTACAGAGTTCTTGAAGTGGTGGC
CTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAGAAAGAGTTGGTAG
CTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAAAAAAGGA
TCTCAAGAAGATCCTTTGATCTTTTACGGGGTCTGACGCTCAGTGGAACGAAACCTACGTTAAGGGATTTTGGTCATGAG
45 ATTATCAAAAAGGATCTTCACTAGATCCTTTTAAATTAATAAATGAAGTTTTTAAATCAATCTAAAGTATATATGAGTAAACTT
GGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCTGTTTATCCATAGTTGCTGACTC
CCCGTCTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCAC
CGGCTCCAGATTTATCAGCAATAAACAGCCAGCCGGAAGGGCCGAGCGAGAAAGTGGTCCGCACTTTATCCGCGCTCCAT
50 CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTGCCAGTTAATAGTTTGGCGAACGTTGTTGCCATTGCTACAG
GCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCACTCAGCTCCGCTTCCCAACGATCAAGGCGAGTTACATGATCCCC
ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGTTGTCAGAAAGTAAGTTGGCCGAGTGTATCACTCATATGGT
TATGGCAGCACTGCATAATTCTTACTGTCTATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGGTACTCAACCAAGTCAT
TCTGAGAATAGTGTATGCGGCGACCGAGTTGCTTCTGCCCGCGCTCAATACGGGATAAATACCGCGCCACATAGCAGAACTTT
55 AAAAGTGCTCATCATTTGGAACGTTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTTCGATGTAA
CCCCTCGTGCAACCCCACTGATCTTACGATCTTTTACTTTCACAGCGTTTCTGGGTGAGCAAAAAACGGAAGGCAAAATG
CCGCAAAAAAGGGAATAAGGGCGACACGGAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTTATCA
60 GGGTTATTGTCTATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGA
AAAGTGCCACCTGACGTC

Ejemplo 5

Se preparó una biblioteca de EBE de *Ralstonia* y vectores de cadena principal que podrían usarse para ensamblar dominios de unión a ADN de *Ralstonia* de longitud completa en cadenas principales de TALEN de *Ralstonia* o *Xanthomonas*, mediante el uso del método de ensamblaje Golden Gate. Los RTN se cotransfectaron en la línea celular C6 de rata y se extrajo el ADN para análisis 48 horas después de la transfección. Un fragmento de ADN de 420 pb que contenía el sitio de unión a RTN se amplificó por PCR. Este amplicón se sometió a continuación al ensayo Cel I mediante el uso del kit de detección de mutaciones Surveyor (Transgenomic) según el protocolo del fabricante. En resumen, el amplicón se desnaturaliza en ADN monocatenario y se rehibrida de nuevo hacia ADN bicatenario. Durante este proceso, si se considera que la agrupación original era una mezcla de secuencias WT y mutadas, habrá hibridación cruzada entre las cadenas WT y mutantes lo que conducirá a la formación de heterodúplex. Tras el tratamiento de esta agrupación rehibridada con nucleasa Surveyor, esta reconoce los heterodúplex y los escinde, lo que genera dos fragmentos más cortos (255 pb y 165 pb) a partir del amplicón original (420 pb).

Terminología:

pRVD: plásmido que contiene un solo EBE de *Ralstonia*. Los EBE individuales se sintetizaron por genes y se clonaron en la cadena principal de subconjuntos XTN de FLASH (XbaI, BamHI).

pFus X: un plásmido del subconjunto que contiene los primeros 10 EBE de cualquier RTN dado. La pieza requerida se sintetizó por genes y se clonó en pHSG-298 (SacI, SbfI).

pFUS Z: un plásmido del subconjunto que contiene 11 EBE hasta el penúltimo EBE de cualquier RTN dado.

Eg: Z4 retiene 11-14 EBE, Z5 retiene 11-15 EBE y Z6 retiene 11-16 EBE. Sintetizado por genes y clonado en pHSG-298 (SacI, SbfI).

XTN-bb: *Cadena principal de TAL de Xanthomonas* que contiene los dominios TAL de *Xanthomonas* N-terminal y C-terminal fusionados con la nucleasa FokI. Esta cadena principal especifica un nucleótido T 5' de la secuencia objetivo especificada por los EBE. Además, contiene la última mitad de EBE que especifica el último nucleótido de la secuencia objetivo. Por lo tanto, existen cuatro plásmidos XTN-bb, cada uno de los cuales especifica un nucleótido final diferente de la secuencia objetivo (igual que las cadenas principales XTN de FLASH).

Todos los plásmidos se almacenan a 150 ng/ul en tampón TE 0,1 x.

Métodos: (construir un dominio de unión a ADN de 16EBE y clonarlo en una cadena principal TALEN de *Xanthomonas*).

El ensamblaje de un constructo efector TALEN o TAL personalizado implica dos etapas: (i) ensamblaje de módulos de repetición (pRVD) en subconjuntos de 1-10 repeticiones y (ii) unión de los subconjuntos en una cadena principal para crear el constructo final.

Construimos un monómero TALEN con un conjunto de 17 RVD (5'-TGATAGTCGCCTTATG-T-3'). Seleccione de los plásmidos pRVD aquellos que codifican los RVD 1-10 en el conjunto mediante el uso de plásmidos numerados en ese orden. Por ejemplo, el plásmido para el primer RVD sería gRTN-1T, el segundo gRTN-2G, el tercer gRTN-3A, etcétera. Los módulos de estos plásmidos se clonarán en el plásmido del subconjunto pFUS-X. A continuación, seleccionar los módulos para los RVD 11-16 en el conjunto de 16 RVD, comenzar nuevamente con los plásmidos numerados a partir de 1. Por lo tanto, para RVD 11 se usaría gRTN-1C, para RVD 12 gRTN-2T, etcétera. Los plásmidos pFUS-Z se numeran 1-10 y deben seleccionarse de acuerdo con el número de EBE que ingresan. Por lo tanto, en nuestro ejemplo, debe usarse pFUS-Z6.

Los pRVD y los plásmidos de subconjuntos (150 ng cada uno) se someten a digestión y unión en una única reacción de 20 ul que contiene 1 ul de Bsal (10 U, New England BioLabs) y 1 ul de ADN Ligasa T4 (2000 U, New England BioLabs) en tampón ADN Ligasa T4. La reacción se incubó en un termociclador durante 10 ciclos de 5 min a 37 °C y 10 min a 16 °C, después se calienta a 50 °C durante 5 min y después a 80 °C durante 5 min. A continuación, se añaden 1 ul de ATP 25 mM y 1 ul de ADNasa Plasmid Safe (10 U, Epicentre). La mezcla se incubó a 37 °C durante 1 h, después se usa para transformar células de *Escherichia coli*. Las células se colocan en placas en agar LB que contiene 50 mg/ml de Kanamicina, durante toda la noche a 37 °C.

Se tamizaron hasta seis colonias de cada transformación con cebadores M13 directos e inversos, a través de PCR de colonias, para identificar clones que contienen un subconjunto de longitud completa. Los clones de subconjuntos pFUS-X de longitud completa deberían producir una banda de 1,1 kb y los clones pFUS-Z6 de longitud completa deberían producir una banda de 700 pb (sumar o restar 105 pb para cada EBE más o menos). Los cultivos se iniciaron durante la noche en cultivos de un pFUS-X de longitud completa y un clon de pFUS-Z6 de longitud completa.

Aislamos ADN plasmídico de sus cultivos pFUS-X y pFUS-Z. Los subconjuntos se unieron en uno de los cuatro plásmidos de cadena principal. Se prepara una mezcla de reacción de digestión y ligadura de 20 ul con 150 ng de cada uno de los plásmidos pFUS-X y pFUS-Z, 150 ng del plásmido de cadena principal, en este caso XTN-bbT, 1ul de Esp3I (10U, Thermo

Scientific) y 1 ul de ADN Ligasa T4 (2000 U, New England Biolabs) en tampón de ADN ligasa T4. La reacción se incubaba después en un termociclador durante 3 ciclos de 10 min a 37 °C y 15 min a 16 °C. La reacción se incubaba después a 37 °C durante 30 minutos adicionales y se calienta a 50 °C durante 5 minutos, después a 80 °C durante 5 minutos. Después de enfriar a temperatura ambiente, se añadieron 1 ul de ATP 25 mM y 1 ul de Safe ADNasa de plásmido (10 U, Epicentre) y se incubaron a 37 °C durante 1 hora. La reacción se usa a continuación para transformar *E. coli* como anteriormente, excepto por la Plasmid Safe. Además, en esta etapa, se usa ampicilina (100 mg/ml) en lugar de Kanamicina para la selección de transformantes.

Tamizamos hasta tres colonias de cada transformación mediante PCR de colonias con cebadores XTN-VF y XTN-VR2 e iniciamos cultivos nocturnos de 1 clon de longitud completa para cada RTN (la banda de 2,1 kb indica el conjunto de 17EBE). A continuación, aislamos el ADN plasmídico e identificamos los clones que contienen el conjunto de repeticiones de longitud completa final mediante secuenciación de ADN con XTN-VF, XTN-VR1 y XTN-VR2.

Cadena principal de subconjunto de XTN digerida con XbaI y BamHI (sitios subrayados):

(BamHI)GGATCCCGGGCCCGCTCGACTGCAGAGGCCTGCATGCAAGCTTGGCGTAA
 TCATGGTCATAGCTGTTTCTCTGTGTGAAATTGTTATCCGCTCACAATTCACACAA
 CATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTA
 ACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTCTG
 TGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGCTATT
 GGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTTCGTTCCGCTGCG
 GCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG
 GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACC
 GTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCA
 TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAG
 ATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGC
 CGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAT
 AGCTCACGCTGTAGGTATCTCAGTTTCGGTGTAGGTTCGTTCCGCTCCAAGCTGGGCT
 GTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAAGTATCG
 TCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT
 AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGG
 TGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGA
 AGCCAGTTACCTTCGGAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCA
 CCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGCAGATTACGCGCAGAAAAAA
 AGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAAC
 GAAAACCTACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCT
 AGATCCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA
 AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATC
 TGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGAT
 ACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAGCCACG
 CTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCG
 CAGAAGTGGTCTCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGG
 GAAGCTAGAGTAAGTAGTTCCGCAAGTTAATAGTTTGCAGAACGTTGTTGCCATTG
 CTACAGGCATCGTGGTGTACGCTCGTTCGTTTGGTATGGCTTCATTCAGCTCCGG
 TTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTT
 AGCTCCTTCGGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGCAAGTGTATCAC
 TCATGGTTATGGCAGCACTGCATAATTCTTCTACTGTATGCCATCCGTAAGATG
 CTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG

5 CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGC
 AGAACTTTAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCTCAA
 GGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
 ATCTTCAGCATCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGG
 CAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCAT
 ACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG
 GATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATT
 10 TCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACC
 TATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTCGGTGATGACG
 GTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAG
 CGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGT
 GTCGGGGCTGGCTTAATACTATGCGGCATCAGAGCAGATTGTAAGTACTGAGAGTGCACC
 15 ATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAATACCGCATCAGGC
 GCCATTTCGCCATTACAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCT
 CTTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTT
 GGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTAAAACGACGGCCAGTGAAT
 20 TCGAGCTCGGTACCTCGCGAATGCATCTAGA(XbaI)

XTN-bb (digerido con BsmBI, los sitios se autoescinden de la cadena principal durante la digestión):

Las secuencias subrayadas se superponen con los subconjuntos pFUS-X y pFUS-Z.

25

XTN-bbA: NNNNNNNNN se reemplaza con TCTAACATC

XTN-bbC: NNNNNNNNN se reemplaza con TCCCACGAC

30

XTN-bbG: NNNNNNNNN se reemplaza con AATAATAAC

XTN-bbT: NNNNNNNNN se reemplaza con TCTAATGGG

35 pFUS-Z se superpone CTGACACCCGAACAGGTGGTTCGCCATTGCTNNNNNNNNNN
 GGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCAATTGTCCAGGCCCGATCCC
 GCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGGAC
 GACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGCTCCCGCATTGATCAA
 40 AAGAACCAACCGGCGGATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCA
 ACTAGTCAAAAGTGAACCTGGAGGAGAAGAAATCTGAACTTCGTGCATAAATTGAA
 ATATGTGCCTCATGAATATATTGAATTAATTGAAATTGCCAGAAATTCCACTCAG
 GATAGAATTCTTGAAATGAAGGTAATGGAATTTTTTATGAAAGTTTATGGATATA
 GAGGTAAACATTTGGGTGGATCAAGGAAACCGGACGGAGCAATTTATACTGTCCG
 45 GATCTCCTATTGATTACGGTGTGATCGTGGATACTAAAGCTTATAGCGGAGGTTA
 TAATCTGCCAATTGGCCAAGCAGATGAAATGCAACGATATGTGGAAGAAAATCA
 AACACGAAACAAACATATCAACCCTAATGAATGGTGGAAAGTCTATCCATCTTCT
 GTAACGGAATTTAAGTTTTTATTTGTGAGTGGTCACTTTAAAGGAACTACAAAG
 50 CTCAGCTTACACGATTAAATCATATCACTAATTGTAATGGAGCTGTTCTTAGTGT
 AGAAGAGCTTTTAATTGGTGGAGAAATGATTAAAGCCGGCACATTAACCTTAGA
 GGAAGTCAGACGGAAATTTAATAACGGCGAGATAAACTTTTAAGGGCCCTTCGA
 AGGTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCAT
 55 CATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTA
 GTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCTTCCTTGACCCTGGAAGGT
 GCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA

60

65

GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGG
 ATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTG
 AGGCGGAAAGAACCAGCTGGGGCTCTAGGGGTATCCCCACGCGCCCTGTAGCG
 5 GCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTG
 CCAGCGCCCTAGCGCCCGCTCCTTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTT
 GCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGCATCCCTTTAGGGTCCGATTTA
 GTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGTGGTTACGTTAG
 10 TGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCT
 TTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTA
 TTCTTTTGATTTATAAGGGATTTTGGGGATTTTCGGCCTATTGGTTAAAAAATGAG
 CTGATTTAACAAAAATTTAACGCGAATTAATTCTGTGGAATGTGTGTCAGTTAGG
 GTGTGGAAAGTCCCCAGGCTCCCCAGGCAGGCAGAAAGTATGCAAAGCATGCATC
 15 TCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAA
 GTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCCGCCCTAACTCC
 GCCCATCCCCGCCCTAACTCCGCCAGTTCCGCCCATTTCTCCGCCCATGGCTGA
 CTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCTGCCTCTGAGCTATTCCA
 20 GAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTCCCGGG
 AGCTTGATATCCATTTTCGGATCTGATCAGCACGTGTGACAATTAATCATCGG
 CATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAAACCATGGCCA
 AGCCTTTGTCTCAAGAAGAATCCACCCTCATTGAAAGAGCAACGGCTACAATCA
 ACAGCATCCCCATCTCTGAAGACTACAGCGTCGCCAGCGCAGCTCTCTCTAGCGA
 25 CGGCCGCATCTTCACTGGTGTCAATGTATATCATTTTACTGGGGGACCTTGTGCA
 GAACTCGTGGTGTCTGGGCACTGCTGCTGCTGCGGCAGCTGGCAACCTGACTTGTA
 TCGTCGCGATCGGAAATGAGAACAGGGGCATCTTGAGCCCCCTGCGGACGGTGTG
 GACAGGTGCTTCTCGATCTGCATCCTGGGATCAAAGCGATAGTGAAGGACAGTG
 ATGGACAGCCGACGGCAGTTGGGATTCGTGAATTGCTGCCCTCTGGTTATGTGTG
 30 GGAGGGCTAAGCACTTCGTGGCCGAGGAGCAGGACTGACACGTGCTACGAGATT
 TCGATTCCACCGCCGCTTCTATGAAAGGTTGGGCTTCGGAATCGTTTCCGGGA
 CGCCGGCTGGATGCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCAC
 CCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAA
 35 ATTTACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTC
 ATCAATGTATCTTATCATGTCTGTATACCGTCGACCTCTAGCTAGAGCTTGGCGT
 AATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGCTCACAATTCACAC
 AACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGC
 40 TAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
 CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTTCGTA
 TTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTGCTTCGGCTG
 CGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCA
 45 GGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAA
 CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAG
 CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAA
 AGATAACAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCC
 TGCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCT
 50 CAATGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTTCGCTCCAAGCTGG
 GCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCGGTAACCTA
 TCGTCTTGAGTCCAACCCGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
 GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAG
 TGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTGGTATCTGCGCTCTGC
 55 TGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAA

CCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGTGCAAGCAGCAGATTACGCGCAGAAA
 AAAAGGATCTCAAGAAGATCCTTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGG
 AACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTC
 5 ACCTAGATCCTTTTAAATTA AAAATGAAGTTTTAAATCAATCTAAAGTATATATG
 AGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGC
 GATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTA
 CGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACC
 10 CACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCG
 AGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTG
 CCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCAGCAAGTTGTTGCC
 ATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTTCAGCTC
 CGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGC
 15 GGTTAGCTCCTTCGGTCCCTCCGATCGTTGTGTCAGAAAGTAAGTTGGCCGCAGTGTTA
 TCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCTATGCCATCCGTAA
 GATGCTTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTAT
 GCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACA
 20 TAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTC
 TCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGACCCCA
 ACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAACAGG
 AAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATAC
 25 TCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG
 AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCGCGCG
 ACATTTCCCCGAAAAGTGCCACCTGACGTC
 GACGGATCGGGAGATCTCCCGATCCCCTATGGTTCGACTCTCAGTACAATCTGCTC
 TGATGCCGCATAGTTAAGCCAGTATCTGCTCCCTGCTTGTGTGTTGGAGGTCGCT
 30 GAGTAGTGCGCGAGCAAAATTTAAGCTACAACAAGGCAAGGCTTGACCGACAAT
 TGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGG
 CCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTAC
 GGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTA
 35 AATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCCATTGACGTCAATAATG
 ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG
 AgTATTTACGGTAACTGCCCACCTTGGCAGTACATCAAGTGTATCATATGCCAAG
 TACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCAG
 40 TACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGC
 TATTACCATGGTGTATGCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTT
 GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT
 GGCACCAAAATCAACGGGACTTTCACAAAATGTCGTAACAACCTCCGCCCCATTGA
 45 CGCAAATGGGCGGTAGGCGGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCT
 GGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTA
 TAGGGAGACCCAAGCTGGCTAGCACCATGGACTACAAAGACCATGACGGTGATT
 ATAAAGATCATGACATCGATTACAAGGATGACGATGACAAGATGGCCCCCAAGA
 AGAAGAGGAAGGTGGGCATTCACCGCGGGGTACCTATGGTGGACTTGAGGACAC
 50 TCGGTTATTTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTCAGGAGCACCG
 TCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGT
 CGCGCTTTCACAGCACCCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAAGAT
 ATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAATTGTAGGGGTCGGTAAA
 55 CAGTGGTCTGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTT
 AGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGA
 GGGGGAGTAACAGCGGTAGAGGCAGTGCACGCCTGGCGCAATGCGCTCACCGGG
 GCCCCCTTGAAC pFUS-X se superpone

60 Fragmentos de pRVD flanqueados por BamHI y XbaI (gen sintetizado, BamHI-EBE-XbaI):
 gXTN-1C:

65

TCTAGAGGTCTCATTGACCCCAGACCAGGTAGTCGCAATCGCGTCAcatgacGGGGG
AAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCA
CGGCAGAGACCGGATCC

gXTN-2C:

TCTAGAGGTCTCACGGCctgactcccgatcaagttgtagcgattgcgtcgCATGACggagggaacaagcattg
gagactgtccaacggctccttcccggtgtgtcaagcccacggAGAGACCGGATCC

gXTN-3C:

TCTAGAGGTCTCAacggTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCcatgatGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGA
TCATGAGAGACCGGATCC

gXTN-4C:

TCTAGAGGTCTCACATGGActgacccagaccaggtagtcgcaatcgcgtaCATGACgggggaaagcaag
ccctgaaaccgtgcaaaggtgtgtgcccgtcctttgtcaagaccacAGAGACCGGATCC

gXTN-5C:

TCTAGAGGTCTCAccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAcatgac
GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAA
GACCAAGAGACCGGATCC

gXTN-6C:

TCTAGAGGTCTCAACCACGGCctgactcccgatcaagttgtagcgattgcgtcgCATGACggagggaaca
agcattggagactgtccaacggctccttcccggtgtgtcaagcccAGAGACCGGATCC

gXTN-7C:

TCTAGAGGTCTCAgcccacggTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCCAT
GATGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATAGAGACCGGATCC

gXTN-8C:

TCTAGAGGTCTCAGGATCATGGActgacccagaccaggtagtcgcaatcgcgtaacatgacgggggaaagc
aagccctggaaccgtgcaaaggtgtgtgcccgtcctttgtcaagaAGAGACCGGATCC

gXTN-9C:

TCTAGAGGTCTCAaagaccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAca
tgacGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCA
AAGAGAGACCGGATCC

gXTN-10C:

TCTAGAGGTCTCACAAGACCACGGCctgactcccgatcaagttgtagcgattgcgtcgcatgacggaggga
acaagcattggagactgtccaacggctccttcccggtgtgtcaagcccTggAAGAGACCGGATCC

gXTN-1T:

TCTAGAGGTCTCATTGACCCCAGACCAGGTAGTCGCAATCGCGTCAAACGGAGG
GGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGA
CCACGGCAGAGACCGGATCC

gXTN-2T:

TCTAGAGGTCTCACGGCctgactcccgatcaagttgtagcgattgcgtcgAACGGTggagggaacaagcattg
gagactgtccaacggctccttcccggtgtgtcaagcccacggAGAGACCGGATCC

gXTN-3T:

5 TCTAGAGGTCTCAacggtTTGACGCCTGCACAAGTGGTCGCCATCGCCTCGAATGGC
GGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAG
GATCATGAGAGACCGGATCC

gXTN-4T:

10 TCTAGAGGTCTCACATGGActgaccccagaccaggtagtcgcaatcgctcaaacggagggggaaagcaagccc
tggaaccgtgcaaaggtgtgtgccgtcctttgtcaagaccacAGAGACCGGATCC

gXTN-5T:

15 TCTAGAGGTCTCAccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAaacgga
GGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAA
GACCAAGAGACCGGATCC

gXTN-6T:

20 TCTAGAGGTCTCAACCACGGCctgactcccgatcaagttgtagcgattgcgtcgAACGGTggagggaaaca
agcattggagactgtccaacggctcctcccgtgtgtcaagcccAGAGACCGGATCC

gXTN-7T:

25 TCTAGAGGTCTCAgcccacggtTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCaatgg
cGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
30 GGATAGAGACCGGATCC

gXTN-8T:

35 TCTAGAGGTCTCAGGATCATGGActgaccccagaccaggtagtcgcaatcgctcaAACGGAagggga
aagcaagccctggaaaccgtgcaaaggtgtgtgccgtcctttgtcaagaAGAGACCGGATCC

gXTN-9T:

40 TCTAGAGGTCTCAaagaccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGTCAA
ACGGAGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTT
GTCAAGAGAGACCGGATCC

gXTN-10T:

45 TCTAGAGGTCTCACAAGACCACGGCctgactcccgatcaagttgtagcgattgcgtccaacggaggggaa
acaagcattggagactgtccaacggctcctcccgtgtgtgtcaagcccaTggAAGAGACCGGATCC

gXTN-1A:

50 TCTAGAGGTCTCATTGACCCCAGACCAGGTAGTCGCAATCGCGTCAaacattGGGGG
AAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCA
CGGCAGAGACCGGATCC

gXTN-2A:

55 TCTAGAGGTCTCACGGCctgactcccgatcaagttgtagcgattgcgtcgaacattggagggaaacaagcattggaga
ctgtccaacggctcctcccgtgtgtgtcaagcccacggAGAGACCGGATCC

gXTN-3A:

60 TCTAGAGGTCTCAacggtTTGACGCCTGCACAAGTGGTCGCCATCGCCAGCaatattGG
CGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCAGGA
65 TCATGAGAGACCGGATCC

gXTN-4A:

5 TCTAGAGGTCTCACATGGActgaccccagaccaggtagtcgcaatcgcgtaAACATTgggggaaagcaag
ccctggaaccgtgcaaaggtgttgccggtcctttgtcaagaccacAGAGACCGGATCC

gXTN-5A:

10 TCTAGAGGTCTCAccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGAACA
TTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTC
AAGACCAAGAGACCGGATCC

gXTN-6A:

15 TCTAGAGGTCTCAACCAAtGGCctgactcccgatcaagttgtagcgattgcgctgaacattggagggaacaagcatt
ggagactgtccaacggctccttcccggtgtgtcaagcccAGAGACCGGATCC

gXTN-7A:

20 TCTAGAGGTCTCAgcccacggtTTGACGCCTGCACAAGTGGTCGCCATCGCCTCCAAT
ATTGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATAGAGACCGGATCC

gXTN-8A:

25 TCTAGAGGTCTCAGGATCATGGActgaccccagaccaggtagtcgcaatcgcgctgaacattgggggaaagc
aagccctggaaccgtgcaaaggtgttgccggtcctttgtcaagaAGAGACCGGATCC

gXTN-9A:

35 TCTAGAGGTCTCAaagaccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGTCGA
ACATTGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTT
GTCAAGAGAGACCGGATCC

gXTN-10A:

40 TCTAGAGGTCTCACAAGACCACGGCctgactcccgatcaagttgtagcgattgcgctgAACATTggagg
gaaacaagcattggagactgtccaacggctccttcccggtgtgtcaagcccaTggAAGAGACCGGATCC

gXTN-1G:

45 TCTAGAGGTCTCATTGACCCCAGACCAGGTAGTCGCAATCGCGaacaataatGGGGGA
AAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTCAAGACCAC
GGCAGAGACCGGATCC

gXTN-2G:

50 TCTAGAGGTCTCACGGCctgactcccgatcaagttgtagcgattgcgaataacaatggagggaacaagcattggag
actgtccaacggctccttcccggtgtgtcaagccacggAGAGACCGGATCC

gXTN-3G:

55 TCTAGAGGTCTCAacggtTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAACAA
CGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGCCA
GGATCATGAGAGACCGGATCC

gXTN-4G:

60 TCTAGAGGTCTCACATGGActgaccccagaccaggtagtcgcaatcgcgaaacaataatgggggaaagcaagccct
ggaaaccgtgcaaaggtgttgccggtcctttgtcaagaccacAGAGACCGGATCC

gXTN-5G:

5 TCTAGAGGTCTCAccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGAACAATA
ATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTTGTC
AAGACCAAGAGACCGGATCC

gXTN-6G:

10 TCTAGAGGTCTCAACCAAtGGCctgactcccgatcaagttgtagcgattgcgaataacaatggagggaacaagcat
tggagactgtccaacggctccttcccggtgtgtgtaagcccAGAGACCGGATCC

gXTN-7G:

15 TCTAGAGGTCTCAgcccacggtTTGACGCCTGCACAAGTGGTCGCCATCGCCAACAAC
AACGGCGGTAAGCAGGCGCTGGAAACAGTACAGCGCCTGCTGCCTGTACTGTGC
CAGGATAGAGACCGGATCC

gXTN-8G:

20 TCTAGAGGTCTCAGGATCATGGActgacccagaccaggtagtcgcaatcggaacaataatgggggaaagc
aagccctggaaaccgtgcaaaggtgtgtccggtccttgtcaagaAGAGACCGGATCC

gXTN-9G:

25 TCTAGAGGTCTCAaagaccacggcCTGACCCCAGACCAGGTAGTCGCAATCGCGAACA
ATAATGGGGGAAAGCAAGCCCTGGAAACCGTGCAAAGGTTGTTGCCGGTCCTTT
GTCAAGAGAGACCGGATCC

gXTN-10G:

35 TCTAGAGGTCTCACAAGACCACGGCctgactcccgatcaagttgtagcgattgcgaataacaatggaggga
caagcattggagactgtccaacggctccttcccggtgtgtgtaagccaTggAAGAGACCGGATCC

Fragmentos de pFUS flanqueados por SbfI y SacI (gen sintetizado, SbfI-pFUS-SacI)

pFUS-X:

40 (SbfI)CCTGCAGGTCGACCGTCTCAGAACTTGAAGAGACCGTACGTGATCGTGGTC
TCATggaTTGAAGAGACG GGTACCGAGCTC(SacI)

pFUS-Z1:

45 CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCACGGCctgaA
GAGACGGGTACCGAGCTC

pFUS-Z2:

50 CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCAacggctgaA
GAGACGGGTACCGAGCTC

pFUS-Z3:

55 CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCACATGGAct
gaAGAGACGGGTACCGAGCTC

pFUS-Z4:

60 CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCAccacggcctga
AGAGACGGGTACCGAGCTC

pFUS-Z5:

65

CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCAACCACGG
CctgaAGAGACGGGTACCGAGCTC

pFUS-Z6:

CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCAgcccacggtct
gaAGAGACGGGTACCGAGCTC

pFUS-Z7:

CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCAGGATCAT
GGActgaAGAGACGGGTACCGAGCTC

pFUS-Z8:

CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCAaagaccacggc
ctgaAGAGACGGGTACCGAGCTC

pFUS-Z9:

CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCACAAGACC
ACGGCctgaAGAGACGGGTACCGAGCTC

pFUS-Z10:

CCTGCAGGTCGACCGTCTCATTGAAGAGACCGTACTGgatcgtGGTCTCATggActgaA
GAGACGGGTACCGAGCTC

Ejemplo 7

EBE consenso de 34aa de metilesterasas y metiltransferasas (nn se reemplaza con RVD relevante): QTTERIVAIGT
nn GGTQALEAVLTALPRVCPGMV

Secuencia de Backtranseq de 34aa QTTERIVAIGT SII GGTQALEAVLTALPRVCPGMV (SH es un RVD no específico)

CAGACCACCGAGAGGATCGTGGCCATCGGCACCAAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG

EBE de Metilesterasa (14 EBE en la cadena principal de XTN):

Fuente azul: EBE de metilesterasa. Todos con RVD SH no específicos en este ejemplo.

Fuente negra: Cadena principal XTN por FLASH.

5

10

15

20

25

30

35

40

45

50

55

60

65

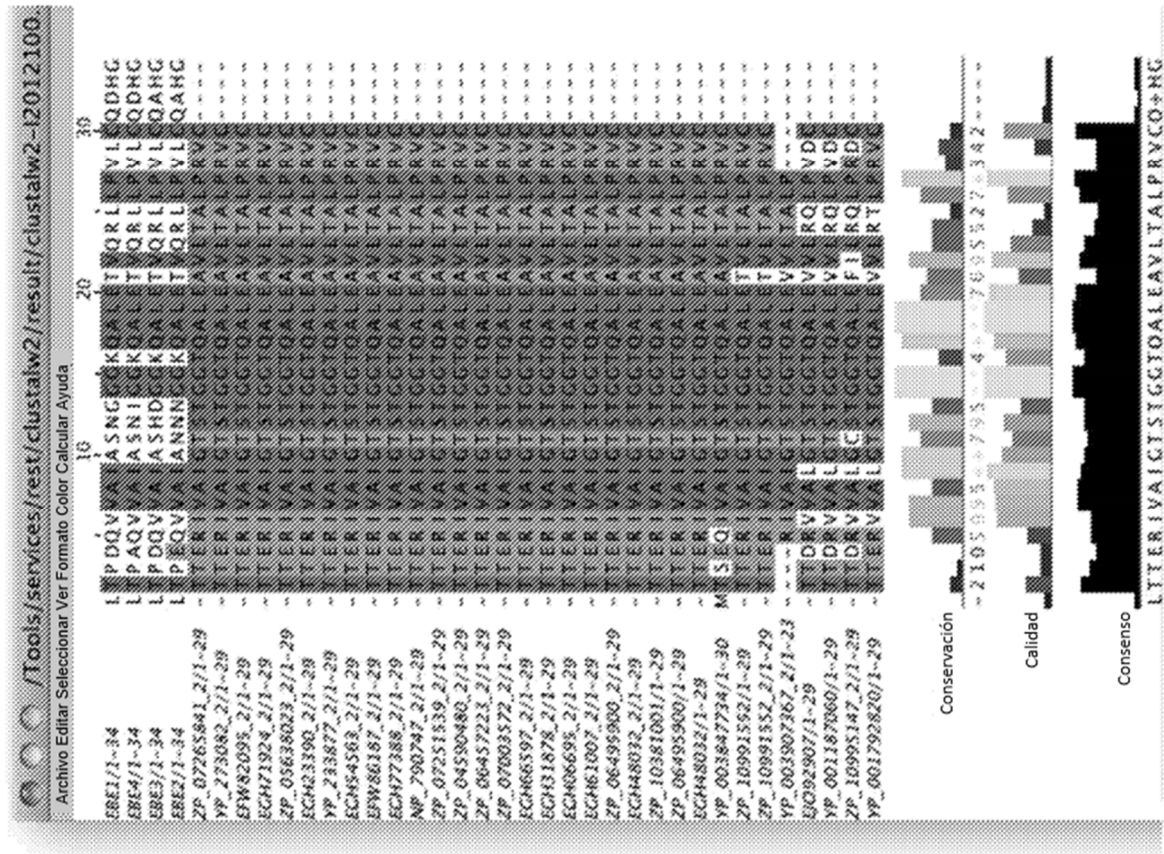
GACGGATCGGGAGATCTCCCGATCCCTATGGTCGACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAG
CCAGTATCTGCTCCCTGCTTGTGTGTGGAGGTGCGTGAGTAGTGC CGGAGCAAAATTTAAGCTACAACAAGG
CAAGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTCGCGATGTACGGG
CCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGC
CCATATATGGAGTTCCGCGTTACATAAECTACGGTAAATGGCCGCGCTGGCTGACCGCCCAACGACCCCGCC
CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGA
gTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCA
ATGACGGTAAATGGCCGCGCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCCTACTTGGCAGTACAT
CTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGATAGCGGTTTG
ACTCACGGGGATTTCGAAGTCTCCACCCCAATTGACGTCAATGGGAGTTTTGTTTTGGCACCAAAATCAACGGGA
CTTTCCAAAATGTCGTAACAACCTCCGCCCAATTGACGCAAAATGGGCGGTAGGCGTGATCGGTGGGAGGTCTAT
ATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGACTCACTATA
GGGAGACCCAAGCTGGCTAGCACCATGGACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGAT
TACAAGGATGACGATGACAAGATGGCCCCAAGAAGAAGAGGAAGGTGGGCATTACCGCGGGGTACCTATG
GTGGACTTGAGGACACTCGGTTATTCGCAACAGCAACAGGAGAAAAATCAAGCCTAAGGTACAGGAGCACCGTC
GCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCATATTGTGCGGCTTTCACAGCACCCGTG
CGGCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGCCACGCACGAGGCAA
TTGTAGGGGTGCGTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGCTTA
GGGGGCTCCGCTCCAGCTCGACACCGGCGAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAG
AGGCAGTGCACGCTGGCGCAATGCGCTCACCGGGGCCCCCTTGAAC

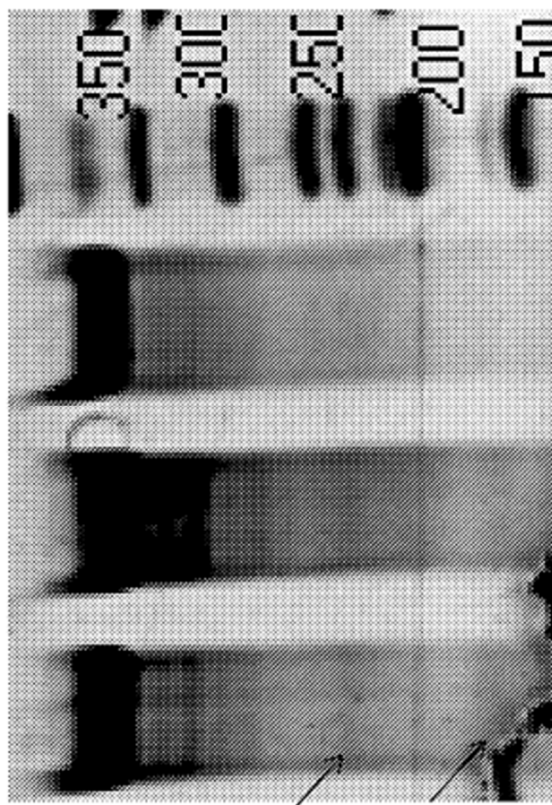
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CAGACCACCGAGAGGATCGTGGCCATCGGCACACAGCCACGGCGGCACCCAGGCCCTGGAGGCCGTGCTGACC
GCCCTGCCCAGGGTGTGCCCCGGCATGGTG
CTGACACCCGAACAGGTGGTCGCCATTGCTAATAATAACGGAGGACGGCCAGCCTTGGAGTCCATCGTAGCCCC
AATTGTCCAGGCCCCGATCCCGCGTTGGCTGCGTTAACGAATGACCATCTGGTGGCGTTGGCATGTCTTGGTGG
ACGACCCGCGCTCGATGCAGTCAAAAAGGGTCTGCCTCATGTCTCCGCATTGATCAAAAGAACCAACCGGCG
GATTCCCGAGAGAACTTCCCATCGAGTCGCGGGATCCCAACTAGTC

REIVINDICACIONES

1. Un polipéptido que comprende
al menos una secuencia de aminoácidos que tiene la secuencia de
LSTEQVVAIASX₁X₂GGKQALEAVKAQLLVLRAPYE (SEQ ID NO: 1), en donde X₁X₂ comprende un díresiduo
variable de repetición (RVD), y en donde el RVD consiste en SI, SN, SH, NP, NH, NT, NK, ND, HN, HY, HD, HH,
RN, RS, NG o GS; y
un dominio efector de ácido nucleico, en donde el dominio efector de ácido nucleico comprende una nucleasa, una
nickasa, un activador de la transcripción, un represor de la transcripción, una metiltransferasa, una desacetilasa o
cualquier fragmento funcional de estos.
2. El polipéptido de conformidad con la reivindicación 1, en donde el polipéptido comprende, además, una segunda
secuencia de aminoácidos que tiene la secuencia de LSTEQVVAIASX₁X₂GGKQALEAVKAQLLVLRAPYE (SEQ
ID NO: 1), en donde X₁X₂ comprende un díresiduo variable repetido (RVD), en donde X₁ es un aminoácido de
origen natural o no natural y en donde X₂ es un aminoácido de origen natural o no natural.
3. El polipéptido de conformidad con la reivindicación 1 o 2, en donde la al menos una secuencia de aminoácidos y/o
la segunda secuencia de aminoácidos comprende la secuencia de aminoácidos de la SEQ ID NO: 1, la SEQ ID
NO: 2, la SEQ ID NO: 3, la SEQ ID NO: 4, la SEQ ID NO: 5, la SEQ ID NO: 6, la SEQ ID NO: 7, la SEQ ID NO: 8,
la SEQ ID NO: 9, la SEQ ID NO: 11, la SEQ ID NO: 12, la SEQ ID NO: 13, la SEQ ID NO: 14, la SEQ ID NO: 15,
la SEQ ID NO: 16, la SEQ ID NO: 17, la SEQ ID NO: 18 o la SEQ ID NO: 19.
4. El polipéptido de conformidad con la reivindicación 1 o 2, en donde el polipéptido no comprende la secuencia de
aminoácidos de la SEQ ID NO: 1, la SEQ ID NO: 2, la SEQ ID NO: 3, la SEQ ID NO: 4, la SEQ ID NO: 5, la SEQ
ID NO: 6, la SEQ ID NO: 7, la SEQ ID NO: 8, la SEQ ID NO: 9, la SEQ ID NO: 11, la SEQ ID NO: 12, la SEQ ID
NO: 13, la SEQ ID NO: 14, la SEQ ID NO: 15, la SEQ ID NO: 16, la SEQ ID NO: 17, la SEQ ID NO: 18 o la SEQ
ID NO: 19.
5. El polipéptido de conformidad con una cualquiera de las reivindicaciones 1-4, en donde el dominio efector de ácido
nucleico comprende una nucleasa.
6. El polipéptido de conformidad con la reivindicación 5, en donde la nucleasa es una endonucleasa.
7. El polipéptido de conformidad con la reivindicación 6, en donde la endonucleasa comprende una endonucleasa
Fok I.
8. El polipéptido de conformidad con la reivindicación 6, en donde la endonucleasa comprende una endonucleasa I-
Sce I.
9. El polipéptido de conformidad con una cualquiera de las reivindicaciones 1-4, en donde el dominio efector de ácido
nucleico comprende una nucleasa de dedos de zinc.
10. Un ácido nucleico que codifica una proteína de conformidad con una cualquiera de las reivindicaciones 1 a 9.
11. Un vector que comprende un ácido nucleico que codifica una proteína de conformidad con una cualquiera de las
reivindicaciones 1 a 9.
12. El vector de conformidad con la reivindicación 11, en donde el vector es un plásmido o una molécula de ARN.
13. El vector de conformidad con la reivindicación 11 o 12, en donde el vector es un retrovirus.
14. El vector de conformidad con la reivindicación 13, en donde el retrovirus comprende repeticiones terminales largas,
una señal de empaquetamiento *psi*, un sitio de clonación y una secuencia que codifica un marcador seleccionable.
15. Una célula que comprende una molécula de ácido nucleico que codifica una proteína de conformidad con una
cualquiera de las reivindicaciones 1 a 9, un ácido nucleico de conformidad con la reivindicación 10 o un vector de
conformidad con una cualquiera de las reivindicaciones 11 a 14.
16. Un kit que comprende: un vector que comprende un ácido nucleico que codifica una proteína de conformidad con
una cualquiera de las reivindicaciones 1 a 9.

Figura 1





Carril 1: muestra transfectada de TALEN de Xanthomonas
Carril 2: muestra transfectada de TALEN de Ralstonia
Carril 3: muestra WT
Carril 4: escalera de ADN

Figura 2